NOT TO BE FURTHER COPIED OR REPRODUCED

(* y VS x *)

12

Looking at Batches of (x,y) Points

Introduction 12-1

12A. Counting-in: lines and polygons 12-2
coordinate, at 12-2
- a striking example 12-11
- drawing smooth ovals 12-14

12B. Middle ovals for schematic plots 12-15
eighths, at 12-16
- the need for arithmetic 12-23

12C. Arithmetic giving approximate polygons and ovals 12-24

12D. Outside and detached points 12-31

12E. Schematic plots 12-36

12F. Further examples 12-40
- several differing batches 12-40
- closely similar batches 12-44
- an economic example 12-47

****

12H. How far have we come? 12-53

****

12K. The third dimension (background) 12-56

****

12L. A word about drawing ovals (warning) 12-57

****

12M. What if we pick less carefully? 12-59

****

12N. Skipping can be used (optional technique) 12-66
- an intermediate form 12-76
coordinatewise skipping 12-76

****

12P Problems for chapter 12 12-79

Exh.		Page
1	A	12-3
2		12-5
3		12-6
4		12-8
5		12-9
6		12-10
7		12-12
8		12-13
9	B	12-18
10		12-19
11		12-20
12		12-21
13	C	12-25/26
14		12-29
15		12-30
16	D	12-34
17		12-35
18	E	12-37
19		12-39
20	F	12-41
21		12-42
22		12-43
23		12-45
24		12-46
25		12-48/5[illegible]
26		12-51
27		12-52
	H	
	K	
28	L	12-58
29	M	12-60
30		12-61
31		12-62
32		12-63
33		12-65
34	N	12-67/68
35		12-69
36		12-70
37		12-71/7[illegible]
38		12-75
39		12-77/78

12

Looking at Batches of (x,y) Points

In dealing with batches of numbers, we found a real place for schematic plots which do not identify all the individual observations. There were two reasons for this:

- the symbol can have greater crispness than the detail, even when the detail is quite clear and viewable.
- as we get more and more observations, the detail becomes more and more confusing, less and less viewable.

When we deal with batches of (x,y) points, each of these considerations still holds, though the batch sizes at which they become relevant -- or controlling -- are somewhat larger.

The need for schematic plotting is at least as great for points as it is for numbers. Indeed, it should be greater, for we can show much about a batch of numbers with a stem-and-leaf display, but we have no comparable tool for looking at batches of (x,y) points.

The methods we are going to suggest are the simplest to describe and use that we know. They assume a willingness to plot a well-chosen expression of one variable against a well-chosen expression of another -- and then take this plot seriously -- seriously enough to make consistent and important use of straight lines to find polygons, inscribed ovals, and the like.

We will combine graphical and arithmetic approaches in the interest of ease of working. We treat small- and moderate-sized batches in this chapter, leaving larger ones to another.

12A. Counting-in: lines and polygons

The basic idea we are going to use is simple:

- plotting a batch of (x,y) points.
- moving a straight boundary inward, while keeping it parallel to its earlier positions.
- counting how many points either have been passed by or lie on the boundary.
- stopping at the exact position where this count reaches a predetermined number.
- drawing in -- physically or mentally -- some or all of the boundary in the position where it stopped.

Exhibit 1 shows various stages of one particular moving boundary and seven points, where the intent is to count in three.

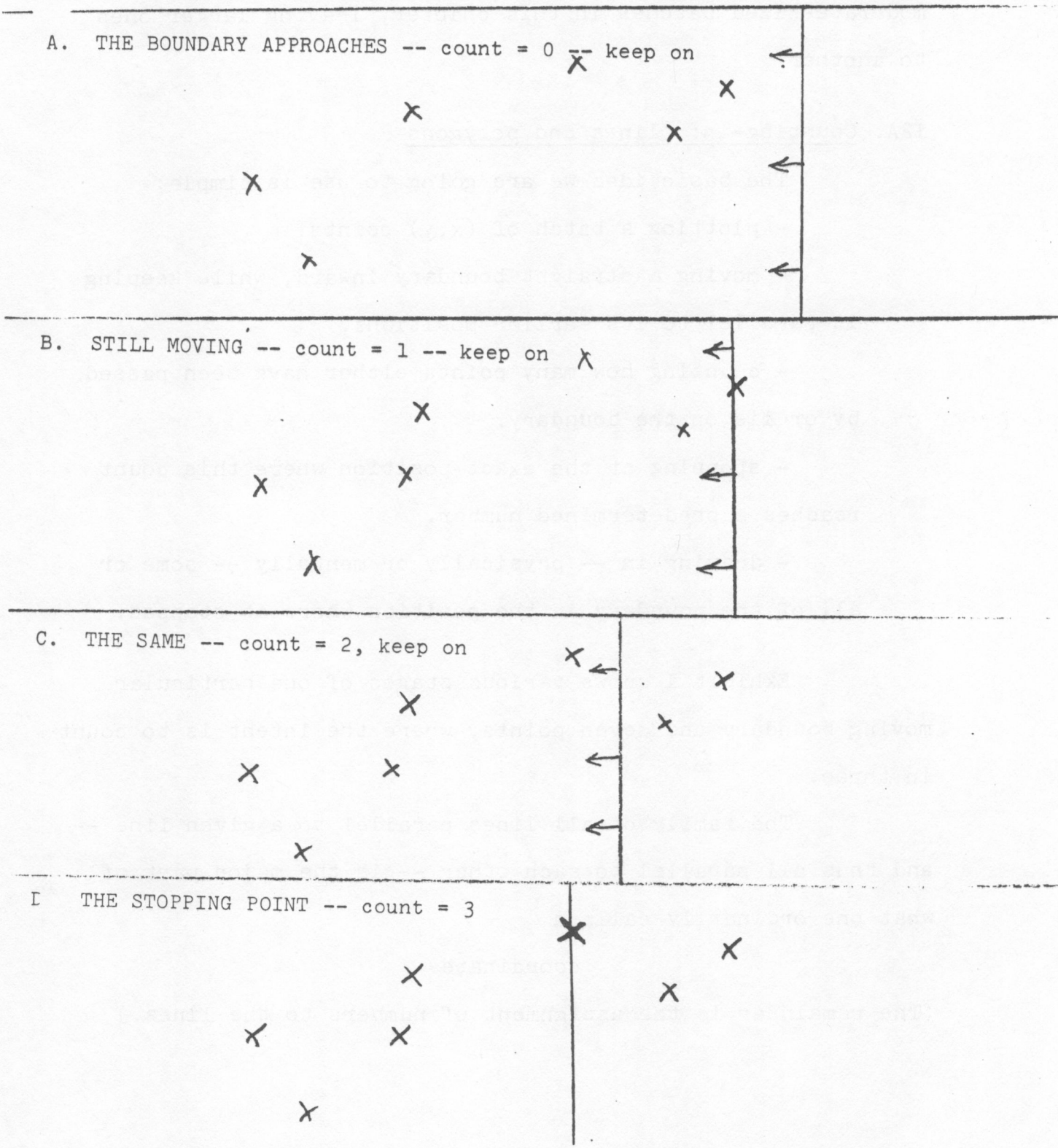
The family of all lines parallel to a given line -- and thus all parallel to each other -- is the major part of what one ordinarily calls a

coordinate

(The remainder is the assignment of numbers to the lines.)

exhibit 1 of chapter 12

Seven Points and a Moving Boundary: Counting in Three



In ordinary graph paper, for example, the vertical lines -- which are spread out horizontally, from line to line -- describe the x-coordinate. Similarly the horizontal lines describe the y-coordinate.

(In more complex graph papers -- those with three or four coordinates, which are made and used for special purposes, and those with two coordinates that are not perpendicular to each other -- this association of a coordinate with a family of parallel lines -- rather than with a scale -- ceases to be unusual and becomes necessary.)

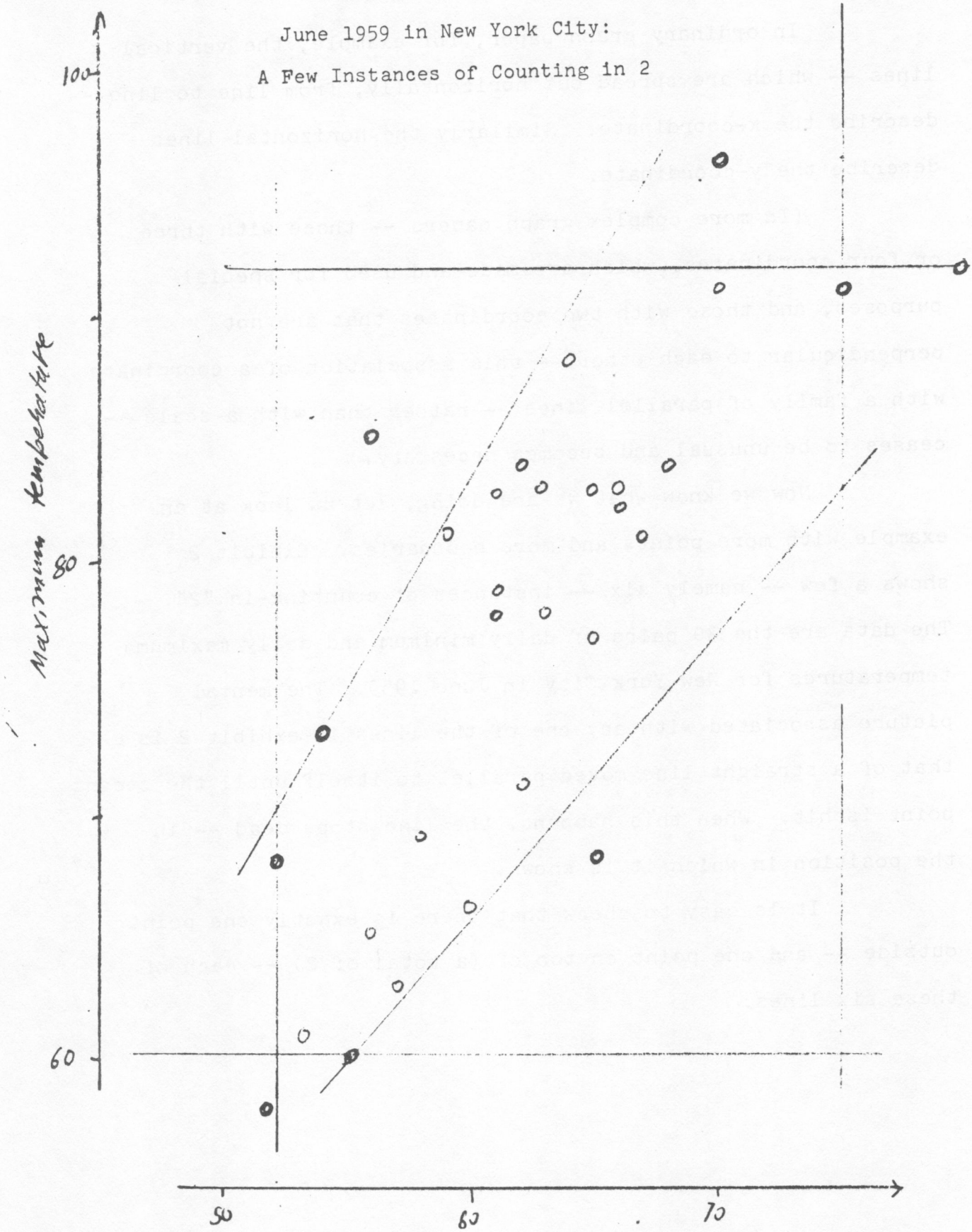
Now we know what we are doing, let us look at an example with more points and more boundaries. Exhibit 2 shows a few -- namely six -- instances of counting-in "2". The data are the 30 pairs of daily minimum and daily maximum temperatures for New York City in June 1959. The mental picture associated with any one of the lines in exhibit 2 is that of a straight line moved parallel to itself until the second point is hit. When this happens, the line stops dead -- in the position in which it is shown.

It is easy to check that there is exactly one point outside -- and one point on top of (a total of 2) -- each of these six lines.

exhibit 2 of chapter 12

June 1959 in New York City:

A Few Instances of Counting in 2



Why stop at 6? Exhibit 3 shows the same data with many "counting-in-2" lines. Clearly these lines do not go everywhere. There is an empty polygon in the middle of the batch. Indeed, giving such a polygon may be a good way for us to tell each other where the "middle of the batch" is. It would be nice to find this polygon by drawing as few lines as we can. Only certain key lines contribute to the final polygon.

When we stop to look and think, we see that the key lines are exactly those where there was a bonus, where the moving line stopped when

two or more

points came on top of it at once, raising the total count from less than two to at least two. Exhibit 4 shows the key lines for our batch of 30 points. There are 11 of them. Exhibit 5 shows the final polygon -- without the excess pieces of lines.

If we are to move toward a schematic plot, we want as simple an object as we dare use. Exhibit 5 takes one step in this direction by also showing a smooth oval which tries to follow the polygon. (In this instance, the oval seems to follow the polygon a little too closely.)

Exhibit 6 shows this counting-in-2 oval, the 30 points, and the reason why we could have expected the oval

12-7

exhibit 3 of chapter 12

June 1959 in New York City:

Many instances of counting in 2

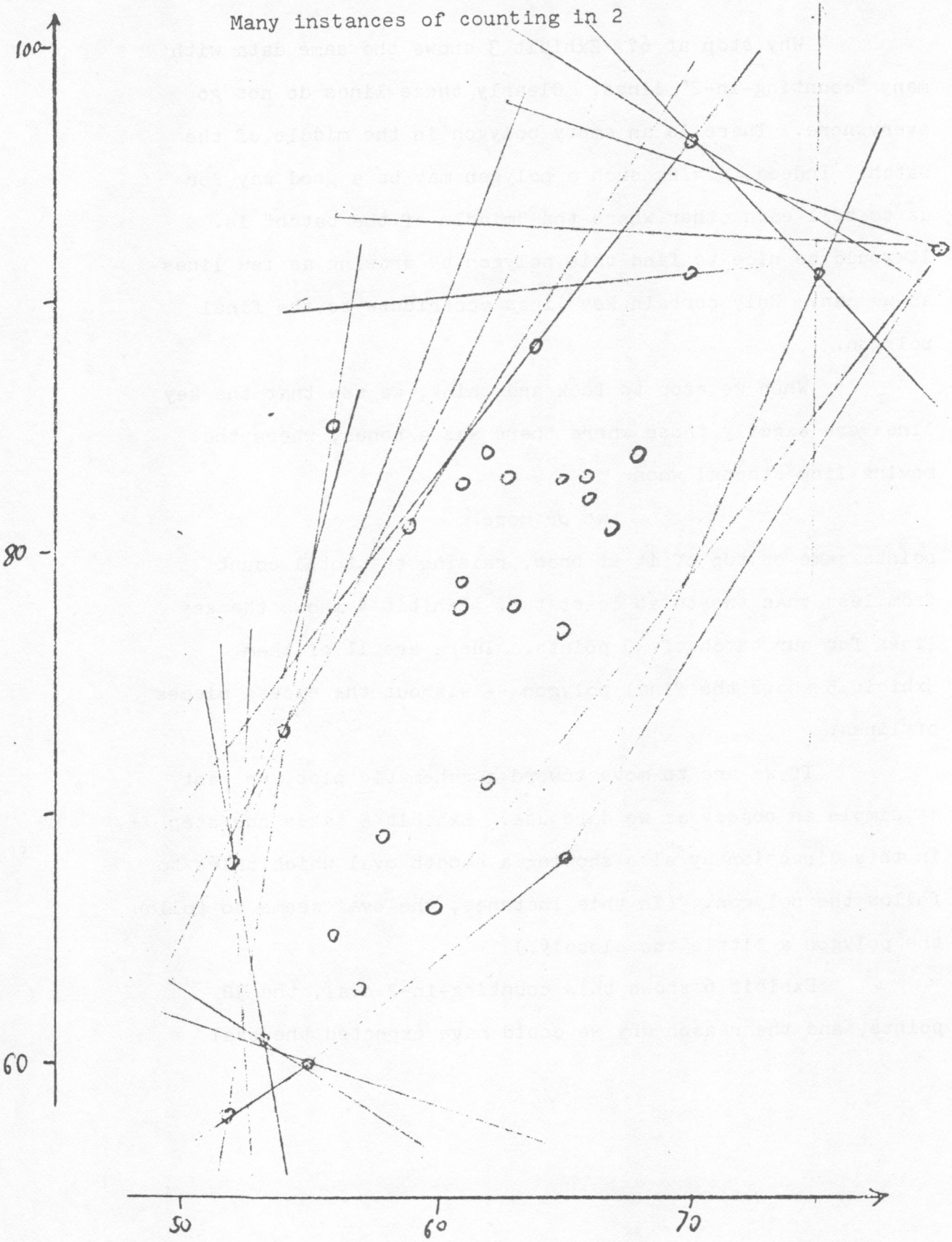


exhibit 4 of chapter 12

June 1959 in New York City:

The Key Instances of Counting in 2

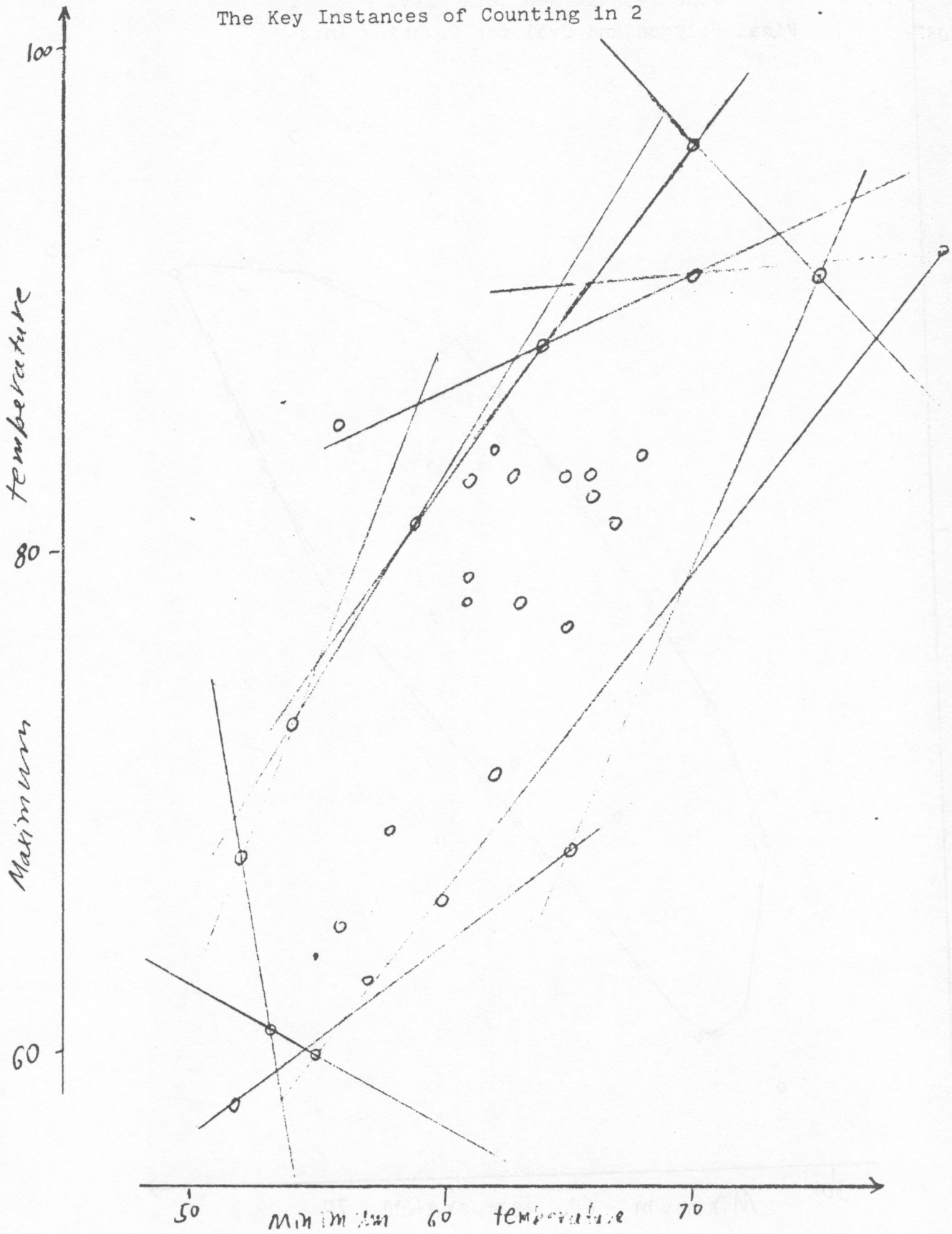


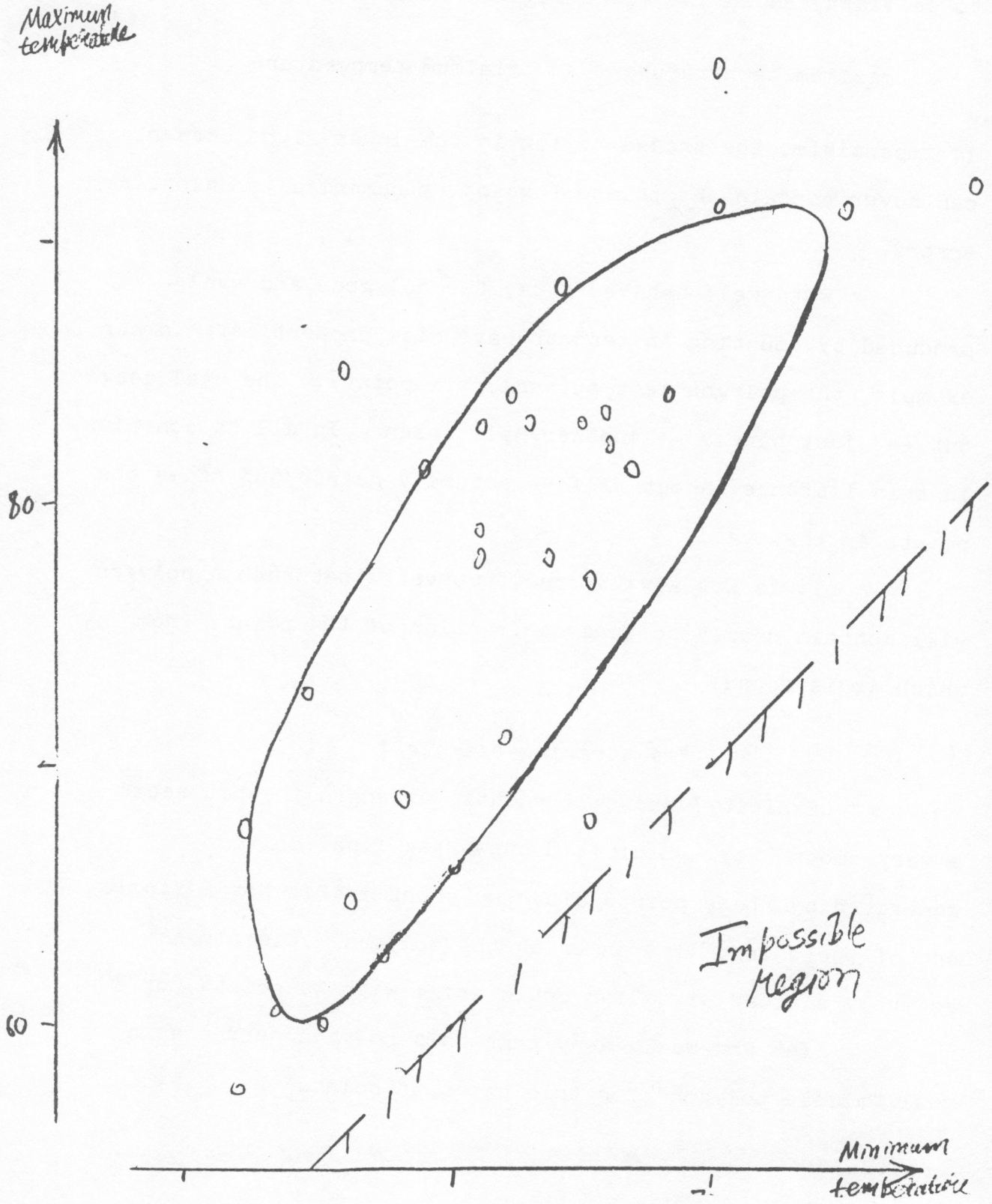
exhibit 5 of chapter 12

June 1959 in New York City:

Final Polygon and Oval for Counting in 2

exhibit 6 of chapter 12

Counting-in-2 Oval, Data Points, and the Impossible Region



to be flattened on the lower right. Since

maximum temperature < minimum temperature

is impossible, the shaded region in the lower right corner can never contain any points (except by numerical or plotting error).

With well-behaved data, the polygons and ovals produced by counting in tend to have nice properties. In our example, the polygon left out only six points. The oval leaves out -- just barely -- another half dozen. In all it contains, in this instance, about half -- actually nearer 60% of -- the points in the oval.

It is not always true, however, that such a polygon will contain nearly so large a fraction of the points from which it is found.

* a striking example *

Exhibit 7 shows 10 points, arranged in what seems a very smooth way. Exhibit 8 shows the final polygon generated by these points after we count in 2. Not a single one of the points lies in or on the polygon. Clearly an entirely similar situation could arise with 30 points, or 300.

Yet who would deny that this polygon serves as a real "middle polygon" for this batch of points?

exhibit 7 of chapter 12

10 Points in a Smooth Pattern

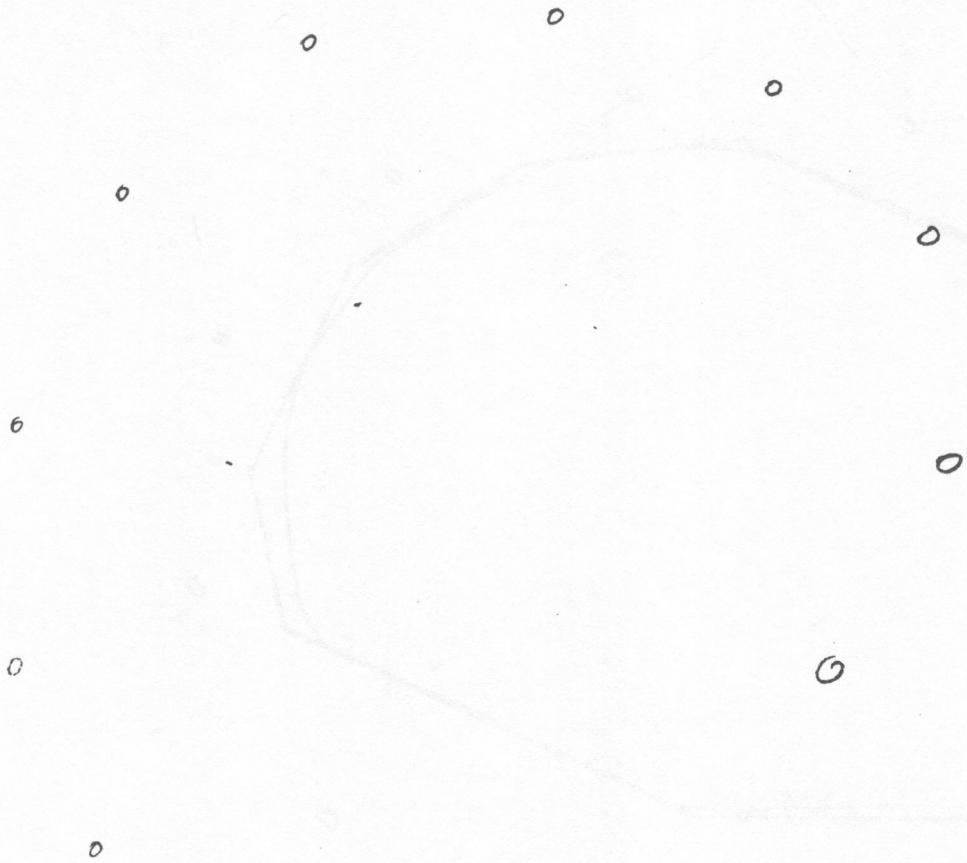
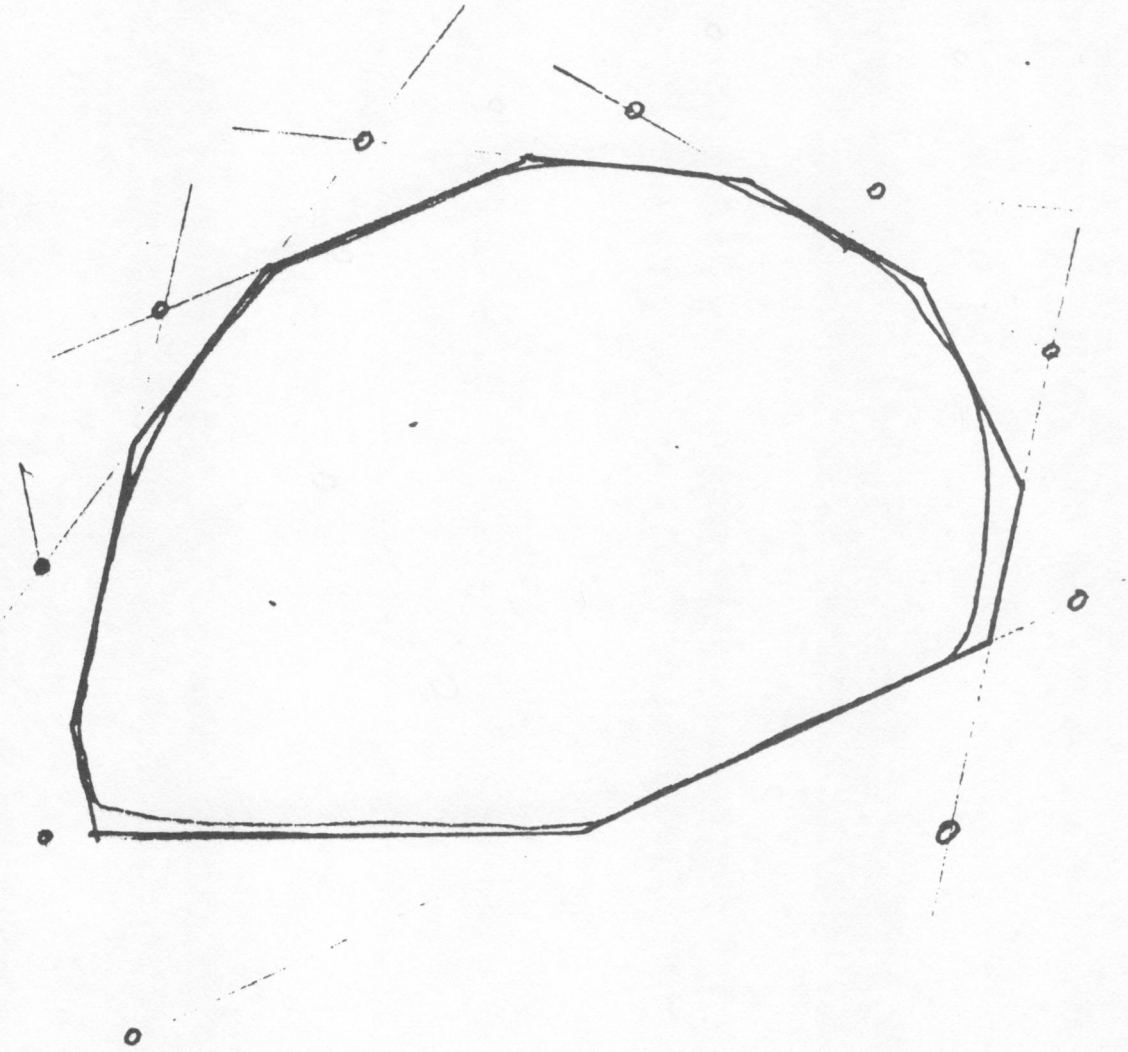


exhibit 8 of chapter 12

The Same Points and the Final Counting-in-2 Polygon

With Its Oval



We can keep alive the spirit of the middle box of chapter 5 -- which showed us the middle of a batch of numbers -- by using polygons defined by counting in. Although these polygons will come close to containing appropriate fractions of the data points for most nicely-behaved batches of points, we cannot expect our middle figures to

always contain

any particular fraction of the points on which they are based. Some batches of points will evade them. Yet middle figures help us even when they are completely evaded.

In particular, an x-and-y-together "midmean" would not always be defined if we tried to define it as the mean of all the points inside some counting-in polygon.

* drawing smooth ovals *

If ovals are to serve us well, we need nice-looking ones. It is easy to have them, if we take the proper steps.

First, we should draw our ovals freehand. (Doing a good job by drawing around "french curves" is difficult, to say the least.)

Next, we always draw with the hand inside the curve, and do not expect to draw more than a quarter oval without turning the paper.

Third, we expect to use two pieces of tracing paper, one to do a little cut and try and the other for the final version.

This last point deserves some expansion. To draw the pictures in exhibit 5, for instance, we began by drawing the polygon on one sheet of tracing paper, call it sheet A. Then we put sheet B on top of sheet A, marked a few points for matching, and tried various slightly different modifications of parts of the desired oval, fitting them into the octagon. Erasures are harmless, and we can even leave incorrect pieces of lines alongside correct ones, since this is not the final version. Once we are satisfied with sheet B, we interchange the sheets, putting sheet A on top and matching up our reference points again. Now we can draw an oval on sheet A by following the oval on sheet B. With practice we find that we only almost follow the old oval and that the new oval is better -- smoother, neater, more visually effective -- than the one on sheet B.

12B. Middle ovals for schematic plots

We now know how to "count-in" and find polygons and ovals. We want to have an oval to play the role of the box of chapter 5. For well-enough behaved data, such a

middle oval

should contain about half the points in the batch.

How far should we count in? When working with batches of numbers, we counted in 1/4 of the way at each end -- and were guaranteed that 1/2 of the numbers in the batch were left in the middle box. We are now working with batches of pairs of numbers. There are many more ways to count in, but it might make sense to count in only half as far in every direction. This would mean counting in about 1/8 the total count. When we try it -- in practice for small and moderate batches, via mathematics for very large ones -- we find counting in 1/8 quite satisfactory. Let us define the

eighths

and the

eighth lines

of a batch of numbers as the result of counting in to a certain number, that number being

$$\text{what } \frac{\text{total count}}{8} \text{ grows to}$$

where "grow to" is defined as in the introduction to chapter 1, in the same sense that hinges are found by counting in what (total count)/4 grows to.

We naturally define the common part of all the strips formed by the pairs of eighth lines as the

middle polygon .

For our previous example, of New York City daily temperatures in spring 1959, we have

$$\frac{30}{8} = 3\frac{3}{4} \quad \text{grows to} \quad 4$$

so we count in 4 along each coordinate. Exhibit 9 shows the key instances of counting in 4. Exhibit 10 shows the middle polygon and middle oval that result.

Exhibits 11 and 12 show two forms of the finished product:

- just an oval -- the analog of what "just a box" would have been in chapter 5.

- an oval and the points -- the analog of the "box-and-dot combined" of chapter 5.

Notice that "just an oval" makes much more sense for batches of (x,y) points than "just a box" would have for batches of numbers. When we were working with batches of numbers, we had an extra dimension to play with. We could use almost any type of schematic plot that could be thought of -- surely it would lie on or near a vertical line! -- and we could stack up as many of these side by side on a piece of paper as we wished. Even when each was made of several elements of different kinds there was no overlap or confusion.

exhibit 9 of chapter 12

June 1959 in New York City:
Key Instances of Counting in 4

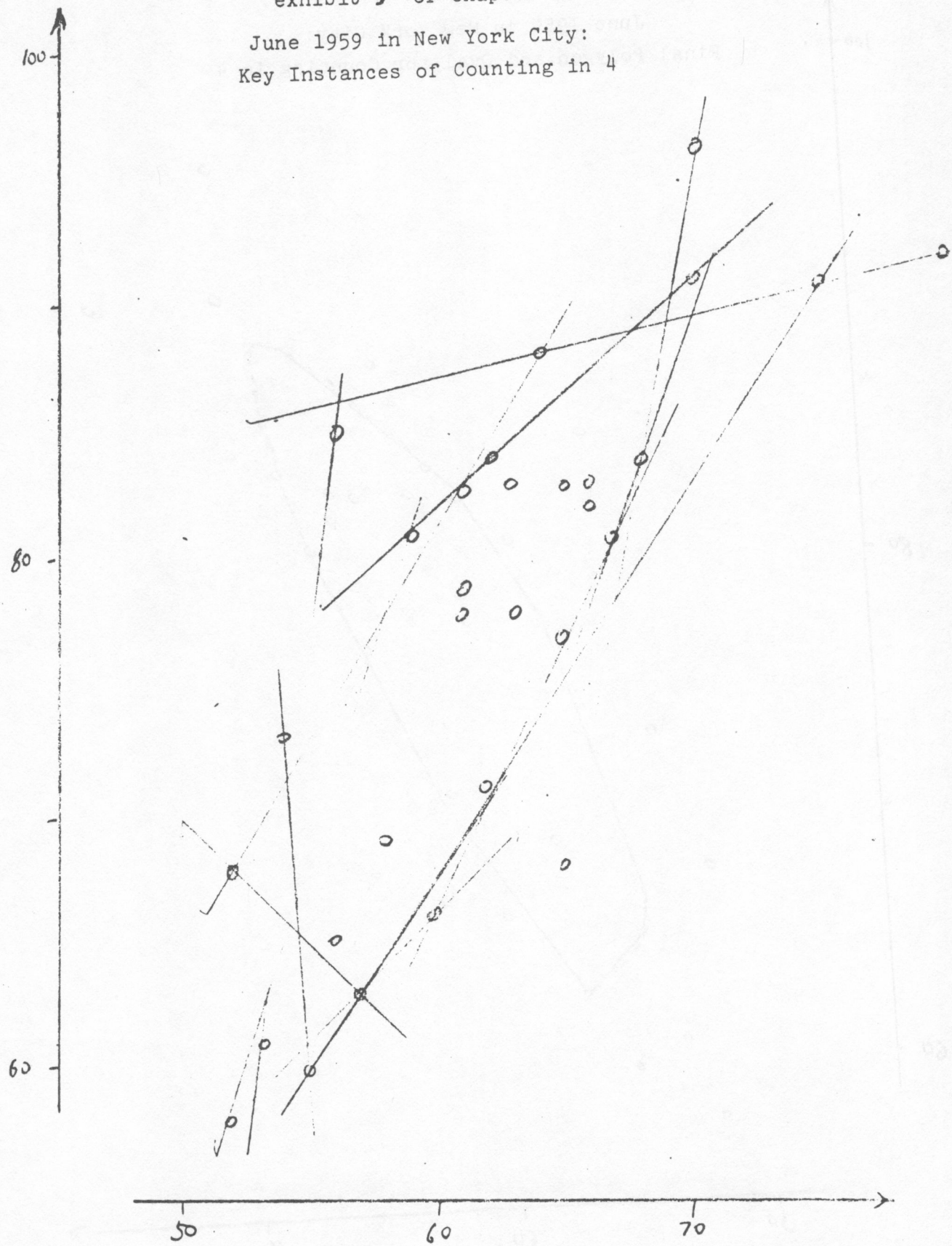
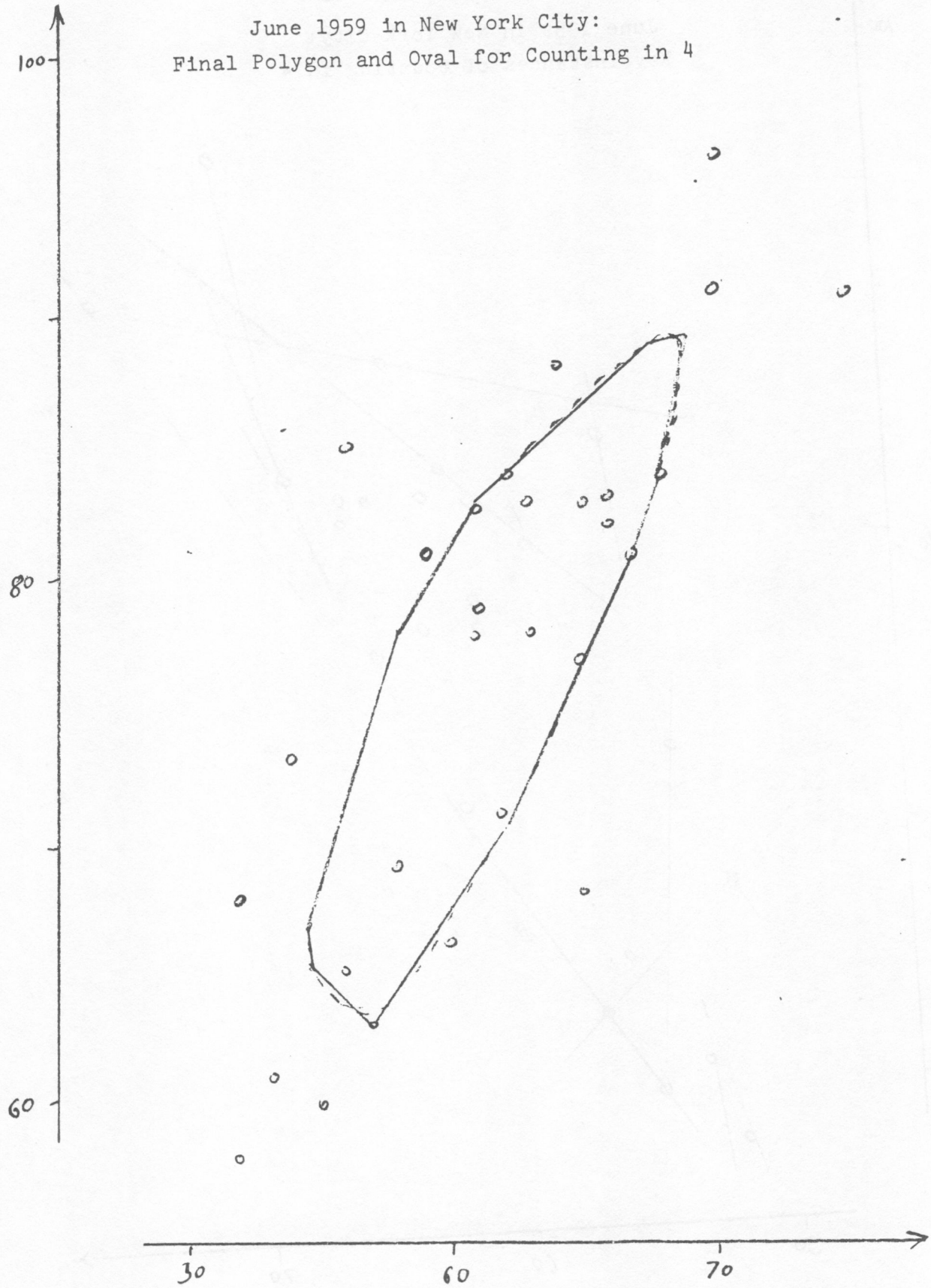


exhibit 10 of chapter 12

June 1959 in New York City:
Final Polygon and Oval for Counting in 4



12-20

exhibit 11 of chapter 12

June 1959 in New York City: Just an Oval

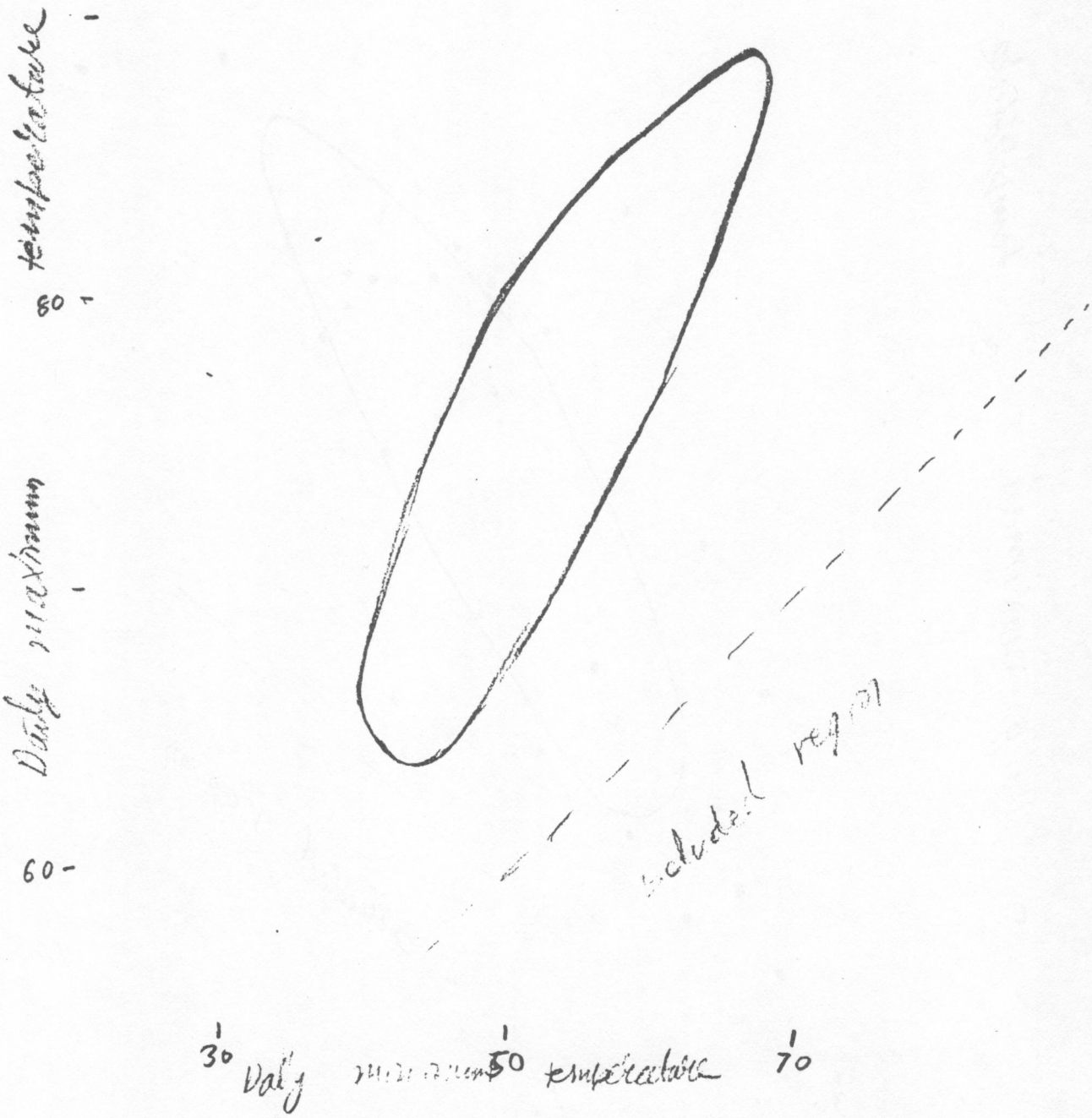
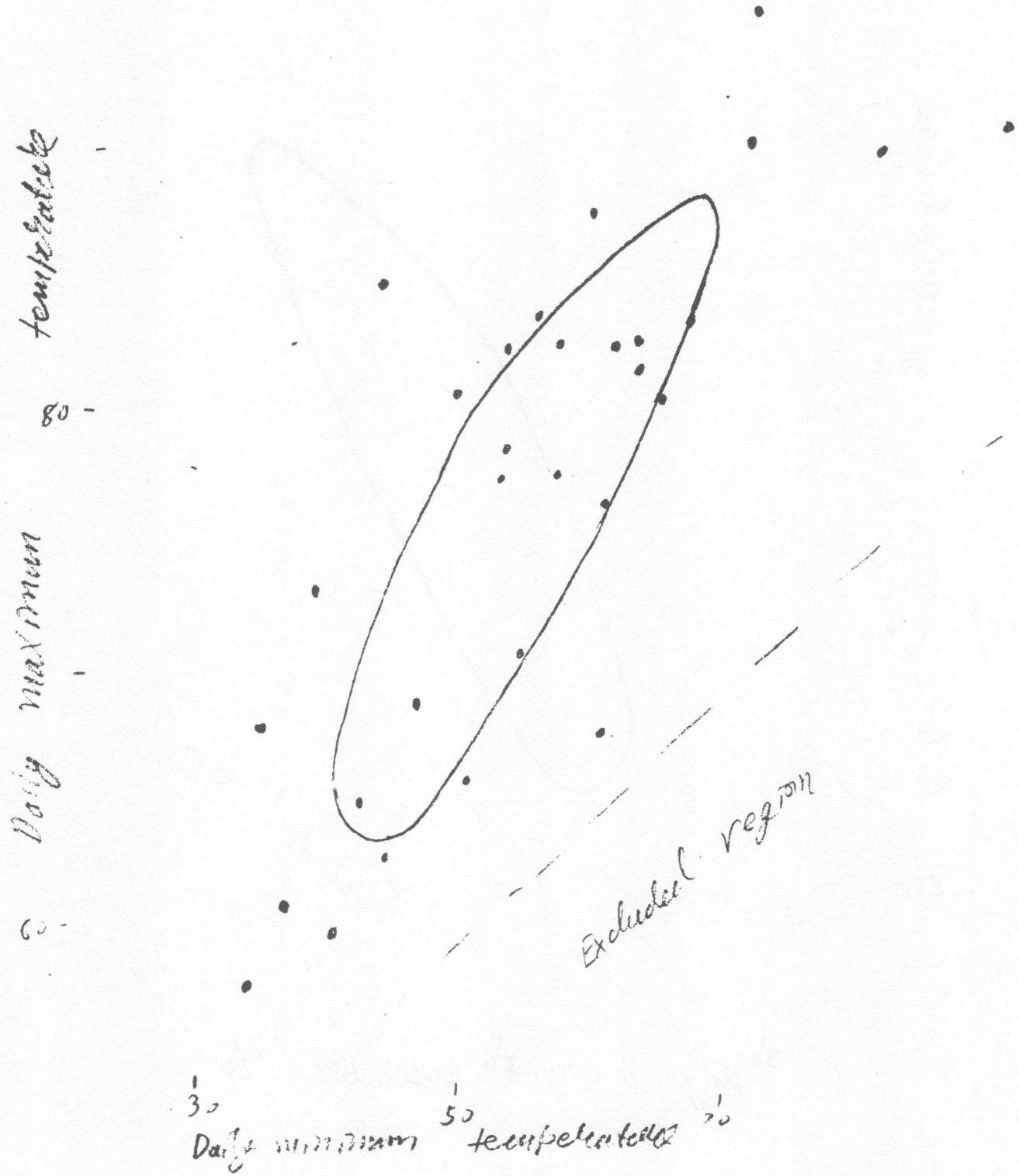


exhibit **12** of chapter 12

June 1959 in New York City:
Oval Combined With Points



Our schematic plots for batches of (x,y) points, however, have to fill out areas to do their task. Two plots that are almost alike, for example, must overlap in a way that cannot help but be confusing. (If we could afford to put each schematic plot on the flat surface of a block of very transparent plastic, we could of course make comparisons by stacking up the blocks, fastening them together, and peering at the pile from many directions. This is not something anyone will do often.) The more complex the structure of our schematic plots, the more confusion.

If two conditions are met, we can do quite well with the "combined approach" of both middle oval and individual points. The two conditions are:

- not too many points -- too many points will ruin the effect.

- no attempt at nearby comparison -- at most comparison of batches far from overlap. Once even the point clouds representing different batches overlap we cannot see them as different entities, even if we try to distinguish them by using different plotting characters for different batches.

* the need for arithmetic *

With 10 or 15 points, the wholly graphical routine of plotting all the points and finding and drawing-in all the counting-in-2 lines is a bearable way to get ready to draw a middle oval. With 100 to 150 points no one of us could stand counting-in 12 to 16 points very long. Hand calculation of all critical lines becomes quite impractical somewhere close to 20 to 30 points. (A handy computer appropriately programmed is -- as always -- a different matter.)

If we are going to make the use of middle ovals bearable, we need some way to approach them that:

- is basically arithmetic, rather than graphical -- it is just too much effort to plot many points.
- does not try to find a complete polygon.

The inevitable price of any such shortcut will be that:

- it reaches only an approximate solution. A good approximation ought not to bother us, since what we want is something to look at.
- performed by different people, the results for a given set of data will differ. This ought not to bother us, either, since drawing a middle oval inside

the middle polygon was already a somewhat personal matter.

There will also be inevitable difficulties with a graphical approach when the scale of our plot has to be small -- perhaps because we are comparing batches that behave quite differently.

12C. Arithmetic giving approximate polygons and ovals

We need a nongraphical approach to middle ovals. We need to do everything we can to streamline this approach, even at the cost of appreciable approximation. If we are willing to count in using only a few coordinates, we can do quite well with a simple arithmetic process.

Exhibit 13 tackles June 1959 in New York City using arithmetic. The first task is collect information about enough counting-in-4 lines so that the oval that corresponds to them will be a reasonable approximation to the oval corresponding to all such lines. In choosing which lines to consider, we needed to be guided by a mixture of easy arithmetic and effective results.

The data will be given to us, ordinarily, in terms of x and y. We should therefore certainly choose to find counting-in lines of the form

$$x = \text{constant}$$

exhibit 13 of chapter 12

Arithmetic Underlying Middle Polygon and Detached Value
for June 1959 in New York City
(x = daily minimum temperature)
(y = daily maximum temperature)
(standard form of calculation)

A) DATA AND COORDINATES

Date	x	y	x+y	x-y	2x-y	3x-y
1	55	60	115	-5	50	105
2	60	66	126	-6	54	114
3	61	79	140	-18	43	104
4	66	82	148	-16	50	116
5	65	77	142	-12	53	118
6	61	83	144	-22	39	100
7	63	83	146	-20	43	106
8	64	88	152	-24	40	104
9	70	91	161	-21	49	119
10	75	91	166	-16	59	134
11	59	81	140	-22	37	96
12	58	69	127	-11	47	105
13	61	78	139	-17	44	105
14	53	61	114	-8	45	98
15	54	73	127	-19	35	89
16	57	63	120	-6	51	108
17	52	58	110	-6	46	98
18	52	68	120	-16	36	88
19	56	65	121	-9	47	103
20	56	85	141	-29	27	83
21	66	83	149	-17	49	115
22	68	84	152	-16	52	120
23	67	81	148	-14	53	120
24	63	78	141	-15	48	111
25	65	68	133	-3	62	127
26	65	83	148	-18	47	112
27	62	84	146	-22	40	102
28	62	71	133	-9	53	115
29	70	96	166	-26	44	114
30	80	92	172	-12	68	148

Notes:

30/8 = 3.75, grows to 4 (eighths)

30/4 = 7.5, grows to 8 (hinges)

(Cont'd)

exhibit 13 of chapter 12 (cont'd.)

B) STEM-AND-LEAVES -- E for eighths, H for hinges

for x

5	3422
5	598766
6	011341322
6	6568755
7	00
7	5
8	0
30	
4E	54
8H	57
8H	66
4E	70

for y

5	8
6	013
6	69858
7	31
7	9788
8	233134134
8	85
9	112
9	6
30	
E	63
H	68
H	84
E	91

for x+y

11	40
11	5
12	001
12	677
13	333
13	9
14	024011
14	869886
15	22
15	
16	1
16	66
17	2
30	
E	120
H	127
H	148
E	161

for x-y

-2	96
-2	204122
-1	8667967658
-1	2142
-0	5686699
-0	3
30	
4E	-22
8H	-20
8H	-09
4E	-06

for 2x-y

2	7
3	
3	9756
4	330404
4	97567987
5	04031233
5	9
6	2
6	8
30	
E	37
H	43
H	52
E	54

for 3x-y

8	3
8	98
9	
9	688
10	40432
10	56558
11	4124
11	68955
12	00
12	7
13	4
13	
14	
14	8
30	
E	96
H	102
H	115
E	120

C) PROBLEM: Carry through panel B for 4x-y.

D) SOURCE: The World Almanac and Book of Facts for 1960, page 450.

and

$$y = \text{constant} .$$

To do this, we need the corresponding eighths. These are found in the first two sections of panel B, which contain first a stem-and-leaf display of the values and then an EHHE listing of eighths and hinges. In particular, then

$$x = 54 \quad \text{and} \quad x = 70$$

are counting-in-4 lines.

What other coordinates should we use besides x and y ? Ease of computation suggests $x+y$ and $x-y$.

Comparative spreads tell us a considerable amount about what others may work well. We have

		for x	for y	ratio
between hinges	dh	9	16	1.8
between eighths	de	16	28	1.8

These ratios correspond to rectangles almost twice as high as they are wide. We would like to try counting-in lines that are somewhere nearly parallel to the diagonal of such a rectangle. Had the height been twice the width,

$$2x-y = \text{constant}$$

and

$$2x+y = \text{constant}$$

would have been such. Thus we may want to try one or both.

A glance at $x+y$ and $x-y$ shows

		for $x+y$	for $x-y$
between hinges	dh	21	11
between eighths	de	41	16

Clearly, as a figure confirms, $x-y$ is much more efficient than $x+y$ in squeezing together. So we try $2x-y$, as a close neighbor.

While we are at it, since we have $x-y$ on one side of $2x-y$, it seems reasonable to add $3x-y$ on the other, as we have done in exhibit 13.

Exhibit 14 shows the working plot coming from exhibit 13. It shows the six pairs of eighth lines, and the corresponding oval, together with the points that lie outside any one of these lines. Exhibit 15 compares the oval thus obtained with the oval from the "final polygon" -- from the middle polygon that is obtained by using all possible eighth lines. The agreement is pleasantly close. So far as finding a middle oval is concerned, our use of only a few selected lines has been quite effective.

What if we had selected fewer or other parts of lines? Some examples are shown in section 12M for those who care to read that far.

exhibit 14 of chapter 12
Working Plot from Exhibit 13

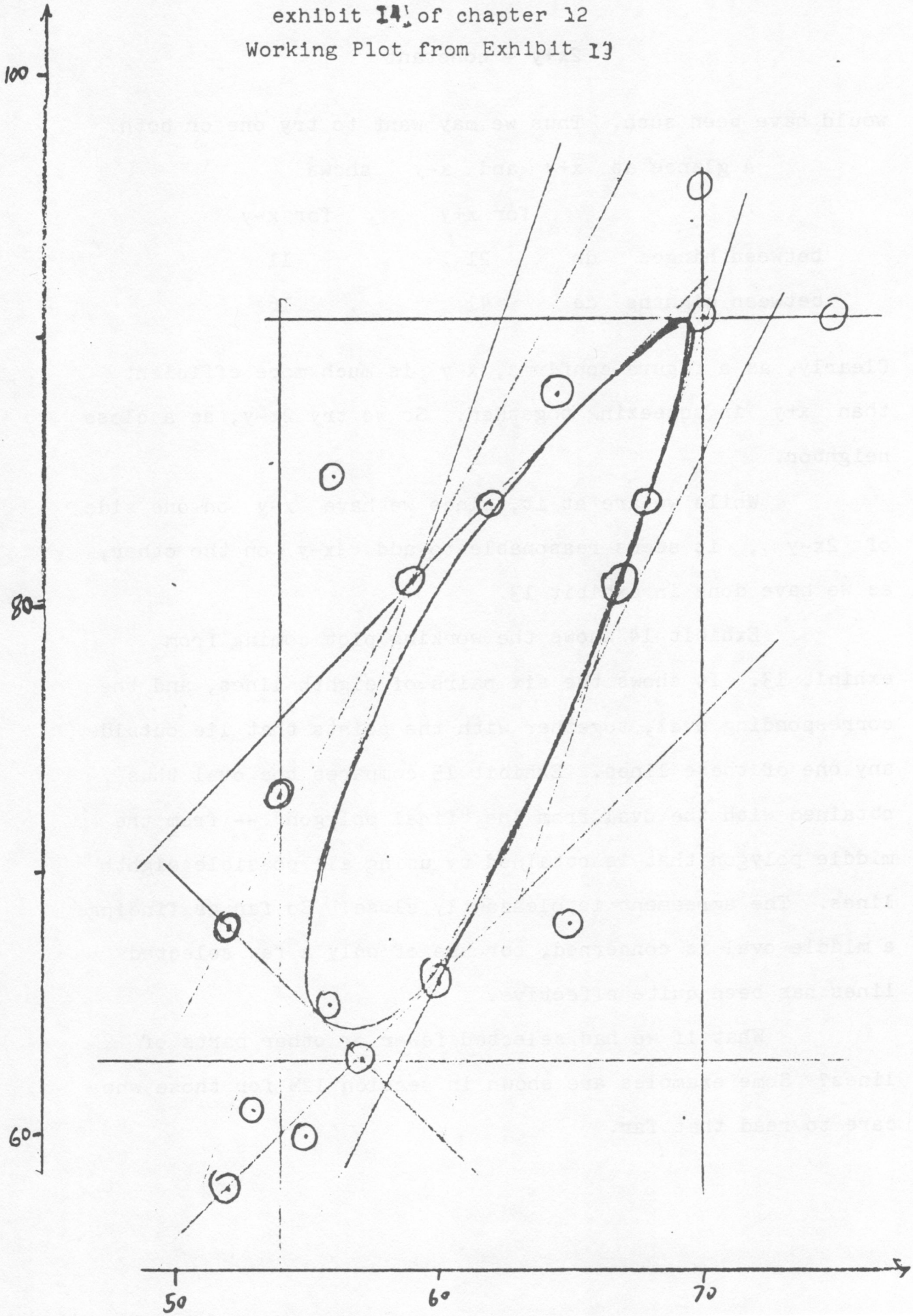
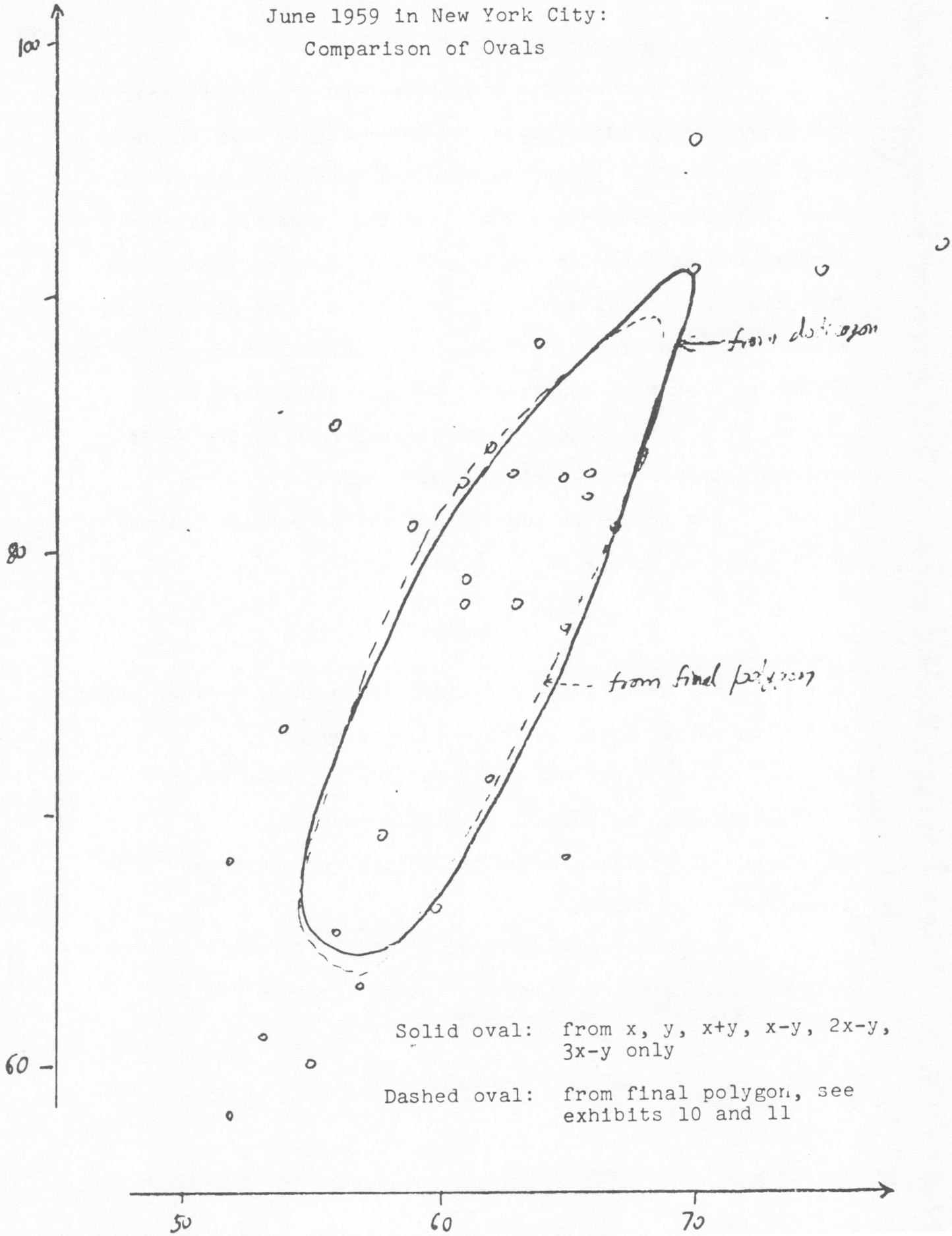


exhibit 15 of chapter 12

June 1959 in New York City:
Comparison of Ovals



Solid oval: from x, y, x+y, x-y, 2x-y, 3x-y only

Dashed oval: from final polygon, see exhibits 10 and 11

12D. Outside and detached points

Now that we have a satisfactory arithmetic process for approaching middle ovals, we are ready to take the next step -- identification of outside and detached points. We hope that a procedure somewhat like that used for batches of numbers can be used for batches of points. To expect to do just the same things would be unrealistic -- we already use eighths instead of hinges. Other things will change too. To expect to work with coordinates -- all coordinates in a utopia, only a few selected coordinates in practice -- is both certainly reasonable and almost necessary.

How are we to identify points as outside points? Which lines do we use as

side lines?

Fortunately the old simple rule applies if we state it in the proper form. We have only to agree to:

- allow one midspread outside whatever cutting points help to define the middle portion

as a means of fixing either side points for numbers or side lines for (x,y) pairs.

In practice we will do this only for those coordinates that we have selected as guides toward a middle oval.

For x in June 1959 -- see exhibit 13 -- we have

$$\begin{array}{l|l} E & 54 \\ H & 57 \\ H & 66 \\ E & 70 \end{array} \quad dh = 9 \qquad \begin{array}{l|l} S & 54-9 = 45 \\ S & 70+9 = 79 \end{array}$$

so that the side lines are x = 45 and x = 79.

We could approach the choice of what is detached as we did in working with numbers, where we used 3/2 of the midspread as our spacing out from the middle-marking points. We can do better -- and keep things simpler -- when working with (x,y) points, by spacing out a distance equal to the spread between eighth points. This could be called an eighth-spread . It will be denoted by

de

for "difference of eighths".

Looking again at x in June 1959 we have

$$\begin{array}{l|l} E & 54 \\ E & 70 \end{array} \quad de = 16 \qquad \begin{array}{l|l} T & 54-16 = 38 \\ T & 70+16 = 86 \end{array}$$

(Incidentally, in this example, using de = 16 as the spacing beyond the eighths gives not too different a result from using

$$\frac{3}{2}\,(dh) = \frac{3}{2}\,9 = 13h \rightarrow 14$$

for this spacing, which leads to 40 and 84 instead of 38 and 86.)

Exhibit 16 takes the results of panel B of exhibit 13, calculates the S- and T-values, and identifies first the outside or detached coordinate values and then the outside or detached points. We find one point detached and two outside.

Notice that:

- all three dates are divisible by 10. Page the numerologist!
- the point for 30th June is shown to be outside by two coordinates and detached by another. Since we use "outside" as a final classification to mean "outside, but not detached", this point is properly listed as detached.

Since we have calculated S- and T-values for each of five coordinates, we can draw 10-sided polygons based on all of

- hinges
- eighths
- S-values
- T-values

Exhibit 17 shows the result. (There need be no concern because some polygons overlap the impossible region . There can properly be directions where points become impossible before they become outside or detached.

exhibit **16** of chapter 12

Arithmetic so far omitted from panel B of exhibit 13
(still June 1959 in New York City)
(standard form of calculation)

A) SUMMARIES from exhibit 13, and the S- and T- VALUES that RESULT -- eighths at 4 and hinges at 8 from 30

	for x		**for y**		**for x+y**	
	T	38 (-)	T	35 (-)	T	79 (-)
	S	45 (-)	S	47 (-)	S	99 (-)
4E	54		E 63		E 120	
8H	57		H 68		H 127	
30 **8**H	66		H 84		H 148	
4E	70		E 91		E 161	
	S	79 (1)	S	107 (-)	S	182 (-)
	T	86 (-)	T	119 (-)	T	202 (-)
dh	9		dh 16		dh 21	
de	16		de 28		de 41	

Examples: 9=66-57, 16=70-54, 45=54-9, 38=54-16, 79=70+9, 86=70+16, etc.

	for x-y		**for 2x-y**		**for 3x-y**	
	T	-38 (-)	T	20 (-)	T	72 (-)
	S	-33 (-)	S	28 (1)	S	83 (-)
4E	-22		E 37		E 96	
8H	-20		H 43		H 102	
30 **8**H	-9		H 52		H 115	
4E	-6		E 54		E 120	
	S	5 (-)	S	63 (1)	S	133 (1)
	T	11 (-)	T	71 (-)	T	144 (1)
dh	11		dh 9		dh 13	
de	16		de 17		de 24	

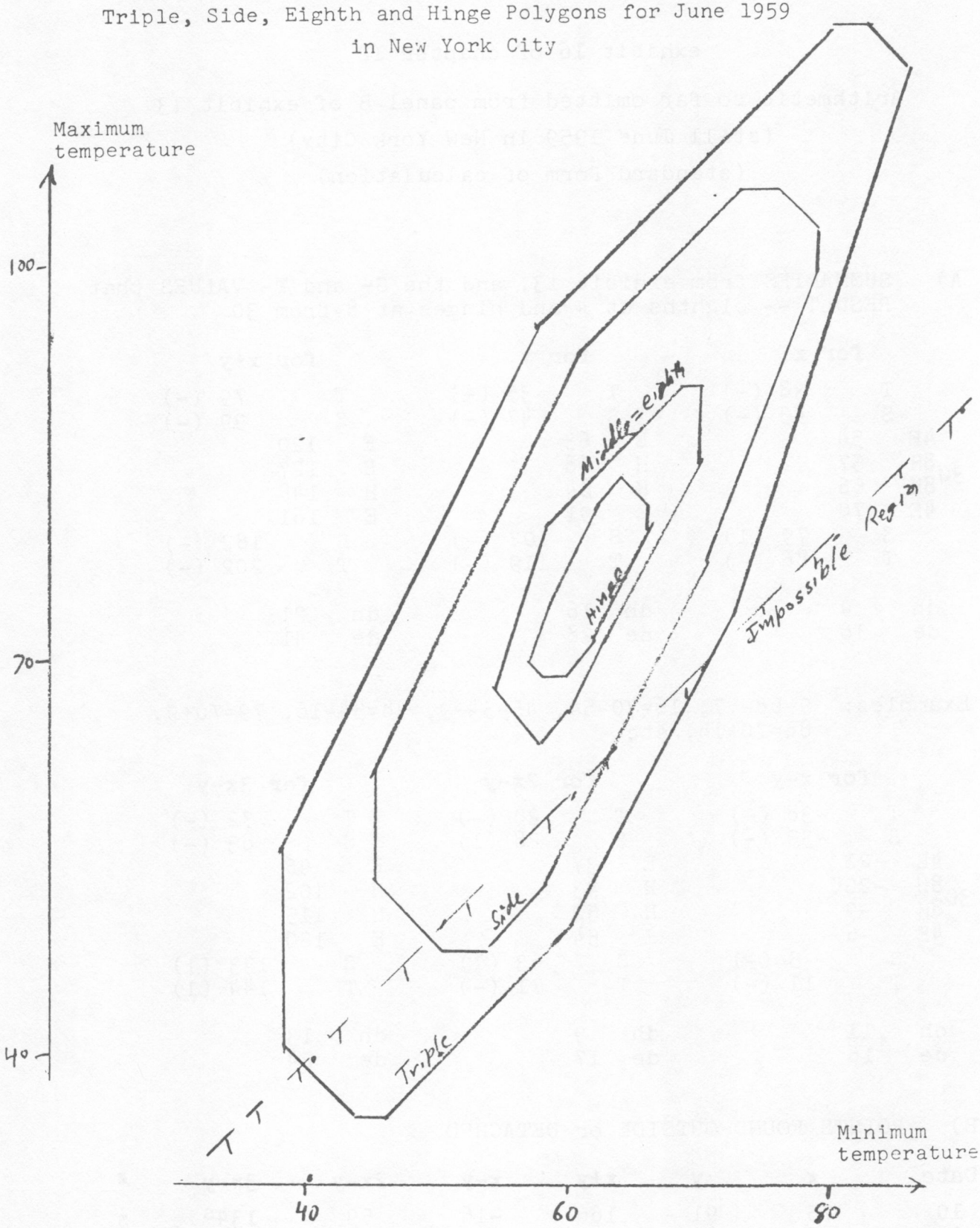
B) POINTS FOUND OUTSIDE or DETACHED

Date	**x**	**y**	**x+y**	**x-y**	**2x-y**	**3x-y**	*
10	75	91	166	-16	59	**134**^{s}	s
20	56	85	141	-29	**27**^{s}	83	s
30	**80**^{s}	92	172	-12	**68**^{s}	**148**^{t}	t

s = outside only
t = detached

exhibit **17** of chapter 12

Triple, Side, Eighth and Hinge Polygons for June 1959 in New York City



12E. Schematic plots

We have already pointed out the greater danger of confusion when we work with schematic plots for batches of points rather than batches of numbers. As we noted, middle ovals alone are often the best choice.

Adding something to the middle oval can sometimes be quite helpful. If we are dealing with a single batch, we can, of course, add any or all other parts more freely than when we are comparing batches. If we try for a complete analog of our schematic plots for batches of numbers, we need to show:

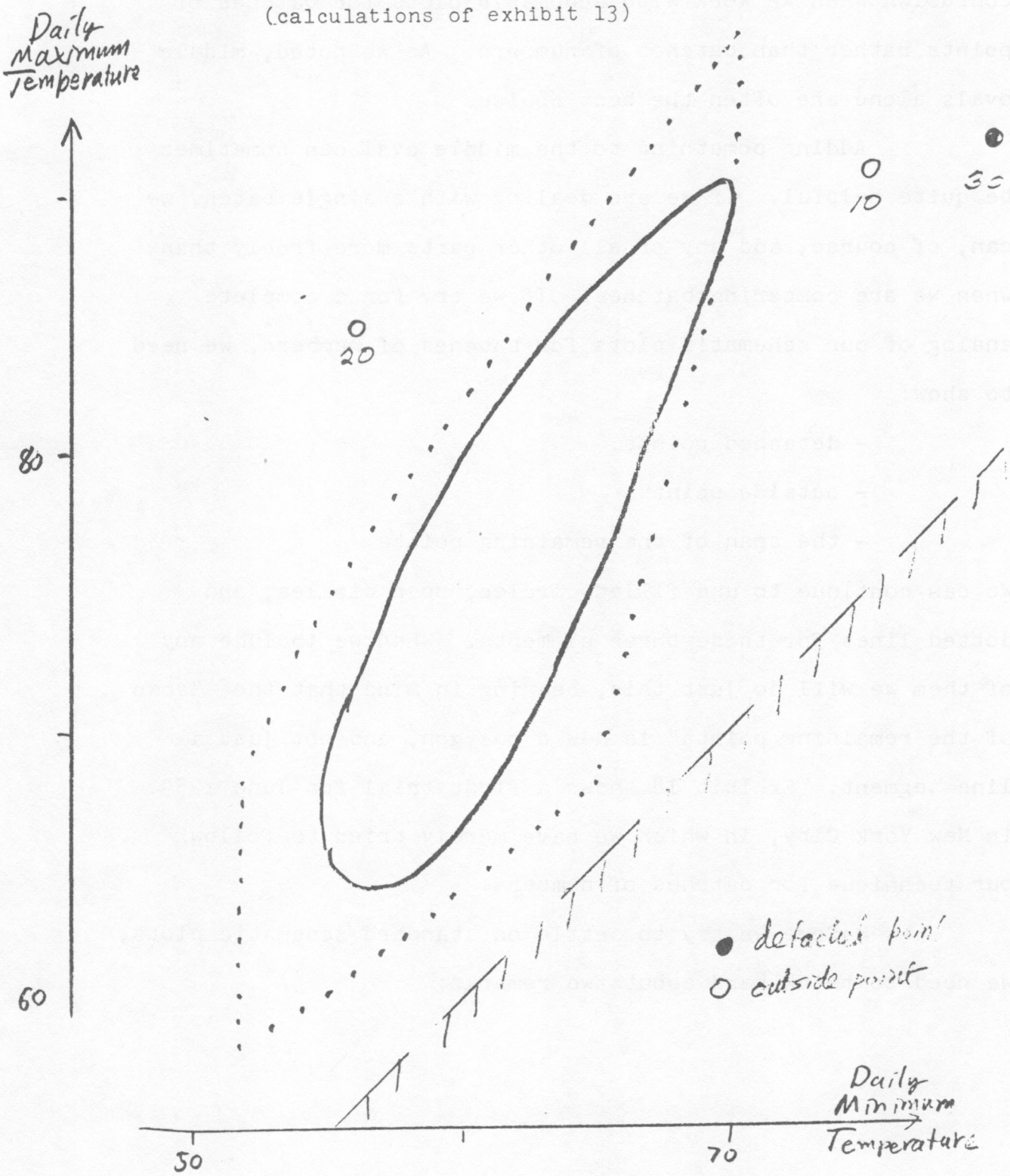
- detached points.
- outside points.
- the span of the remaining points.

We can continue to use filled circles, open circles, and dotted lines for these three elements. When we include any of them we will do just this, bearing in mind that the "span of the remaining points" is now a polygon, and not just a line segment. Exhibit 18 shows a first trial for June 1959 in New York City, in which we have merely tried to follow our technique for batches of numbers.

Before we try to settle on standard schematic plots, we need to think hard about two remarks:

exhibit 18 of chapter 12

June 1959 in New York City
Complete Schematic Plot
First Try, Using chapter 5
(compare exhibits 15 and 19)
(calculations of exhibit 13)



- to avoid confusion in situations where two or more batches of (x,y) points are being compared, what we add to the middle oval should be connected -- probably as a whole, at least in large pieces.

- if we connect a large piece to a very stray point, the line making the connection will seize more of our attention than it deserves.

We have done moderately well with the first of these remarks. The nature of the "span of the remaining points" led us naturally to a dotted polygon. What we have now to consider is attaching to this polygon some or all of the points outside it. The second remark warns us of the need for care. We do not want to go too far with our attachments.

This, of course, is why we have called some points

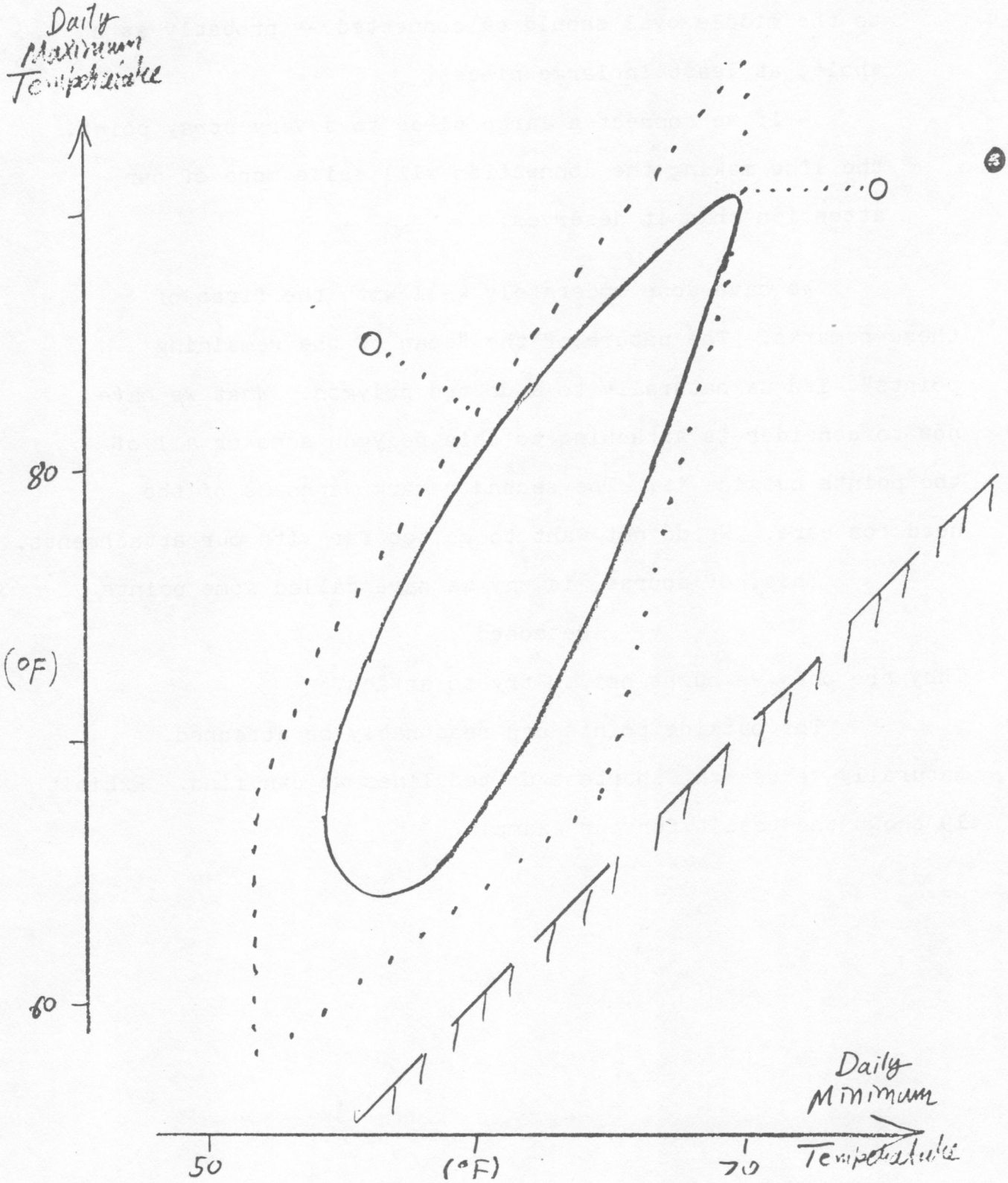
detached .

They are ones we ought not to try to attach.

The outside points can reasonably be attached. Naturally we use the shortest dotted lines we can find. Exhibit 19 shows the result for our example.

exhibit 19 of chapter 12

June 1959 in New York City
Complete Schematic Plot in Standard Form
(compare exhibits 15 and 25)



We are now well fitted up for a variety of schematic plots. The degree of comparison and overlap will determine which one serves us best in a particular situation. Exhibit 20 lays out a natural sequence of six choices from those most subject to confusion to those least subject to it. We can -- and should -- use whichever type of schematic plot serves us best.

Together with the points-and-oval plot these choices cover most of our needs.

12F. Further examples

It is now time to look at a few further examples.

* several differing batches *

It is easy to begin with New York City temperatures for other months in the first half of 1959. Exhibit 21 shows the middle ovals for all six months from January to June.

Clearly we dare not add more detail to this picture. If we did, we would see only a mess. (Here is a place where we wish we had full three-dimensional plotting and vision. If we had, we could put each month on a separate parallel plane and make good use of more complete schematic plots.)

Exhibit 22 shows the result of taking three spaced months -- February, April, and June -- and attempting a plot including remaining polygons. The result is moderately satisfactory.

exhibit 20 of chapter 12

A Sequence of Types of Schematic Plot

Middle oval	Remaining polygon	Outside points	Detached points	(*)
Solid	Dotted	Separate	Solid	Most[1]
Solid	Dotted	Attached[2]	Solid	
Solid	Dotted	Attached	Tagged[3]	
Solid	Dotted	Attached	(none)	
Solid	Dotted	(none)	(none)	
Solid	(none)	(none)	(none)	Least[1]

[1] "Most" and "least" apply both to how subject the plot is to confusion and to how detailed an impression of the batch the plot gives.

[2] Attached by a dashed line.

[3] Plotted with a character keyed to the rest of the plot for that batch. (This is a different thing from identifying the state or date involved.)

exhibit **21** of chapter 12

Middle Ovals for Six Months: 1959 in New York City

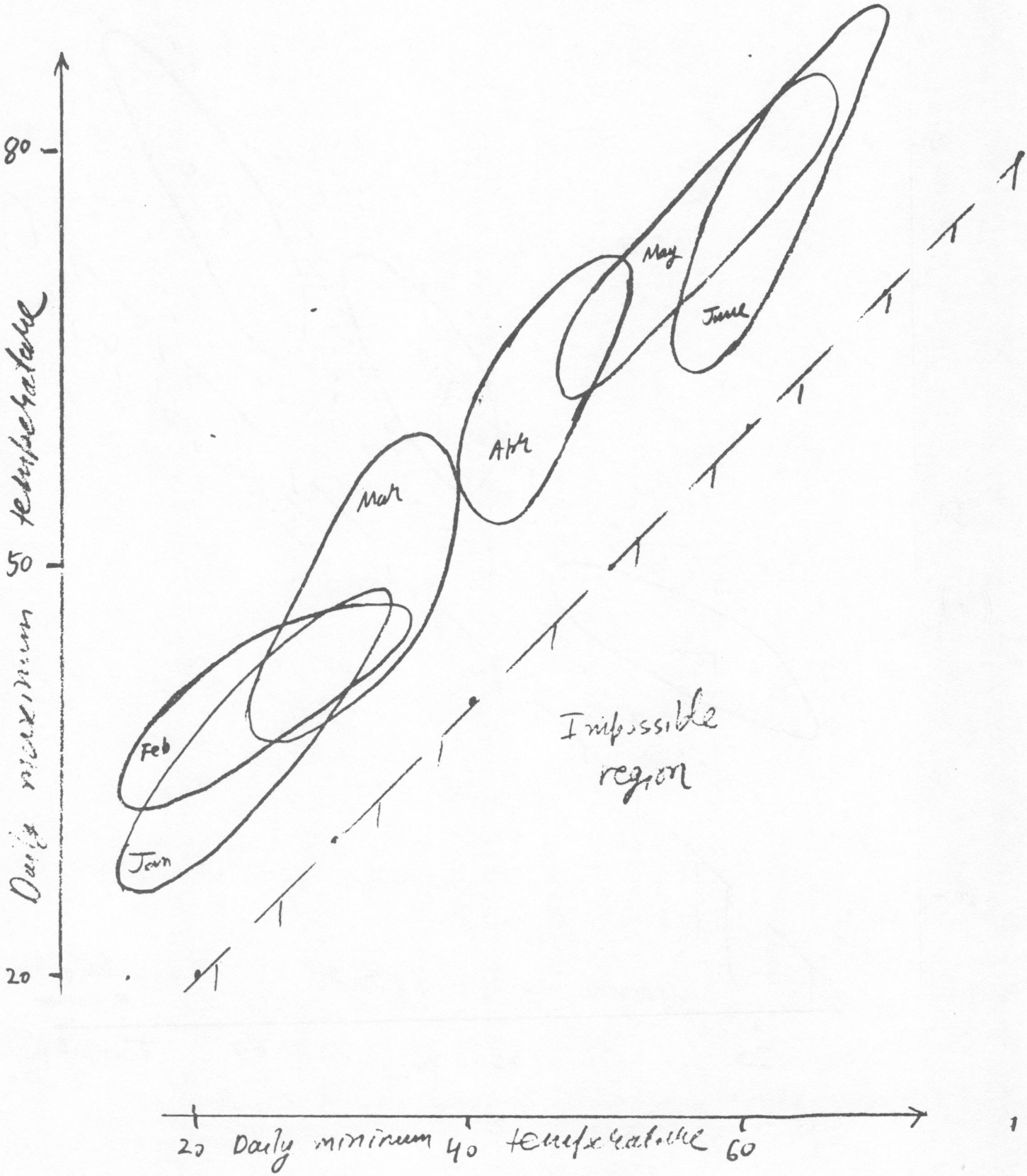
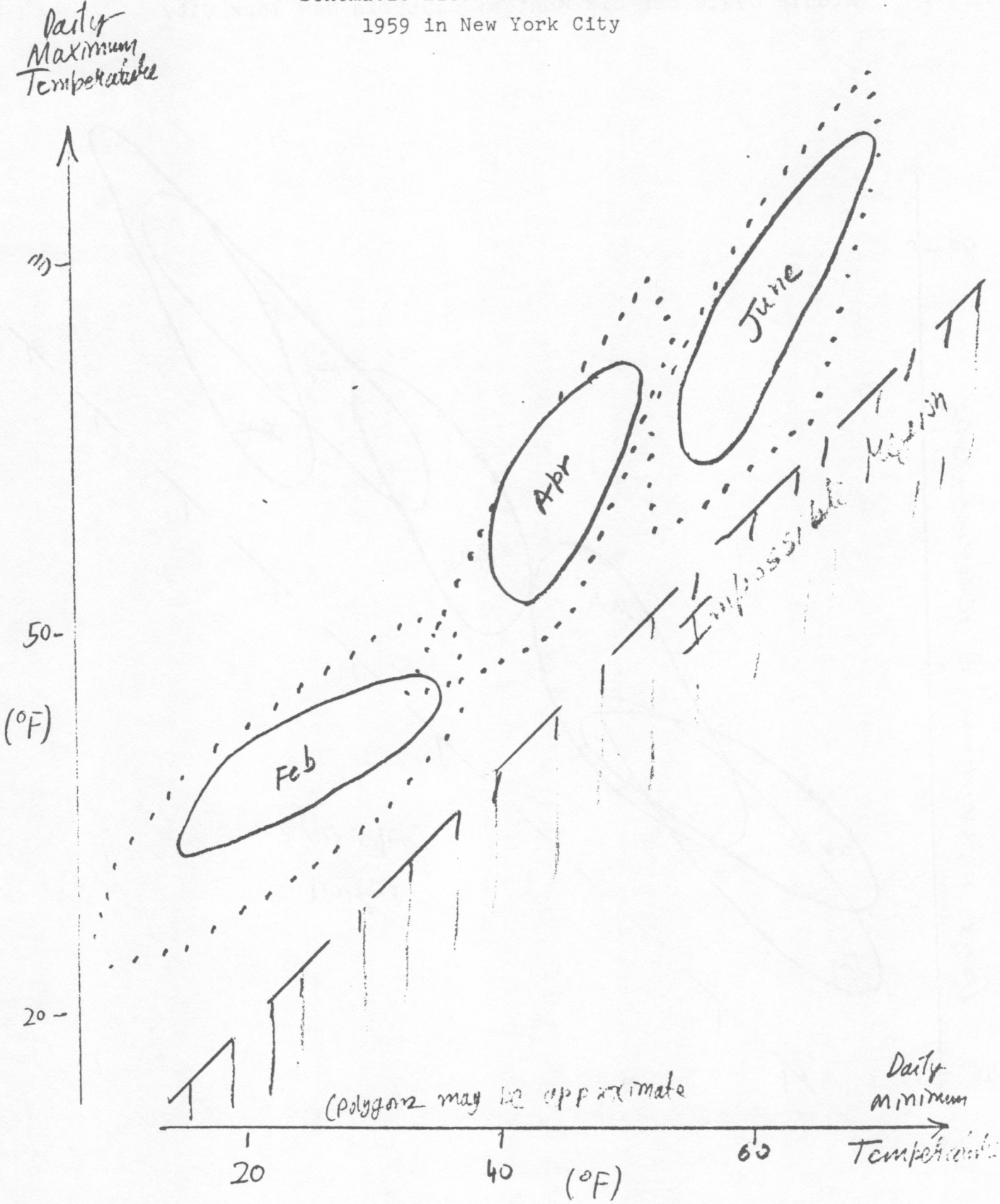


exhibit 22 of chapter 12
Schematic Plots for Three Months
1959 in New York City



If we were to omit the batch for April, we could clearly go ahead with complete schematic plots, including both outside and detached points, for February and June in the same picture.

* closely similar batches *

Comparing two batches that are quite similar is a rather different problem. With luck -- surprisingly enough -- we can show both middle ovals and both remaining polygons.

Exhibit 23 shows daily minimum and maximum temperatures for New York City in June 1958. Exhibit 24 compares two Junes -- 1958 and 1959. The differences are striking. Clearly 1959 tended to:

- have much hotter high maximum temperatures
- have also cooler low maximum temperatures
- have somewhat hotter high minimum temperatures
- show a somewhat greater increase in maximum temperatures for a given change in minimum temperature.

Notice that -- because they are dotted rather than dashed -- the remaining polygons contribute to the comparison rather than hindering it. (In particular, the last of the four points above would have been put somewhat more strongly if we had had only middle ovals to compare.)

exhibit 23 of chapter 12

A Year Earlier

June 1958 in New York City

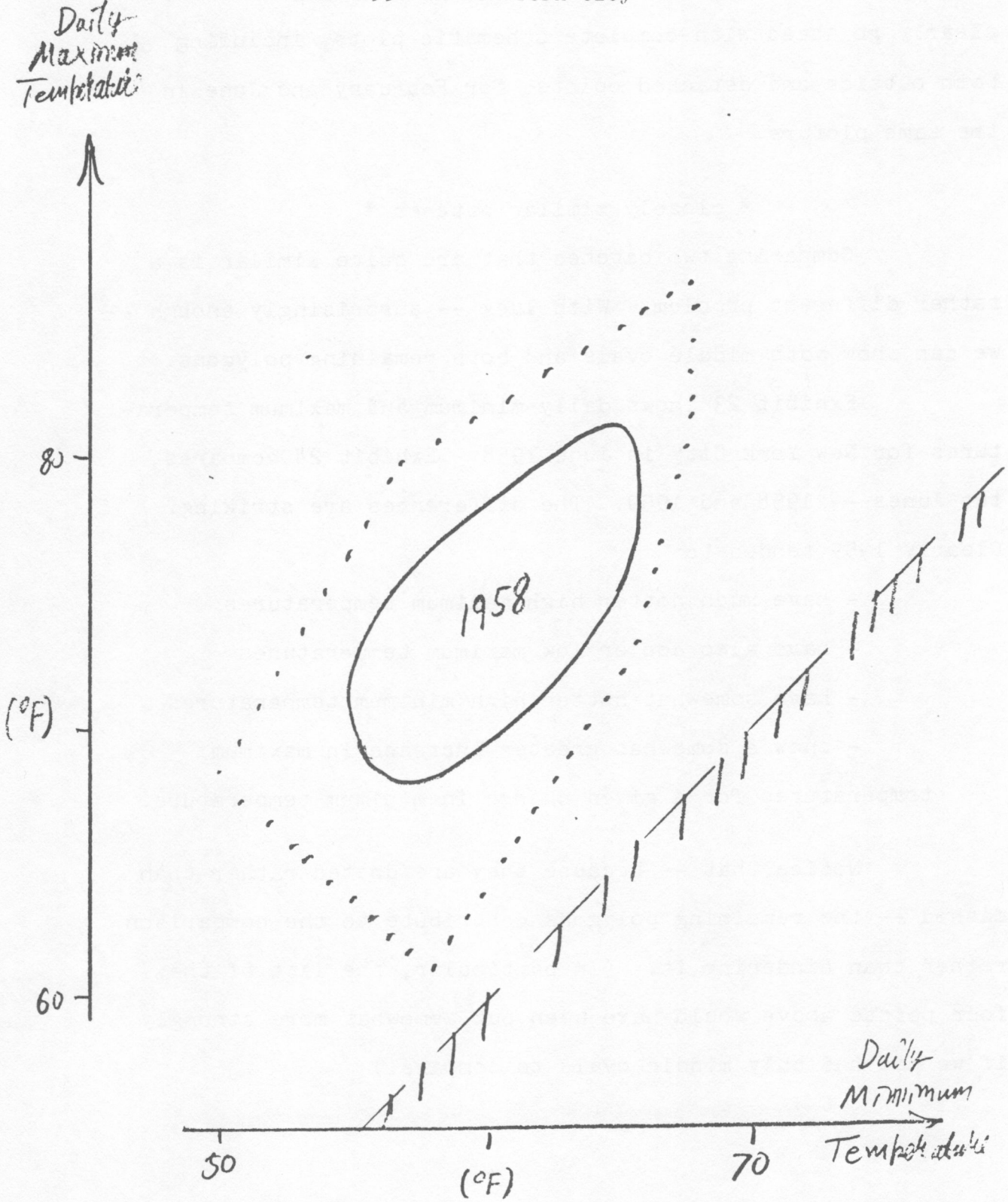
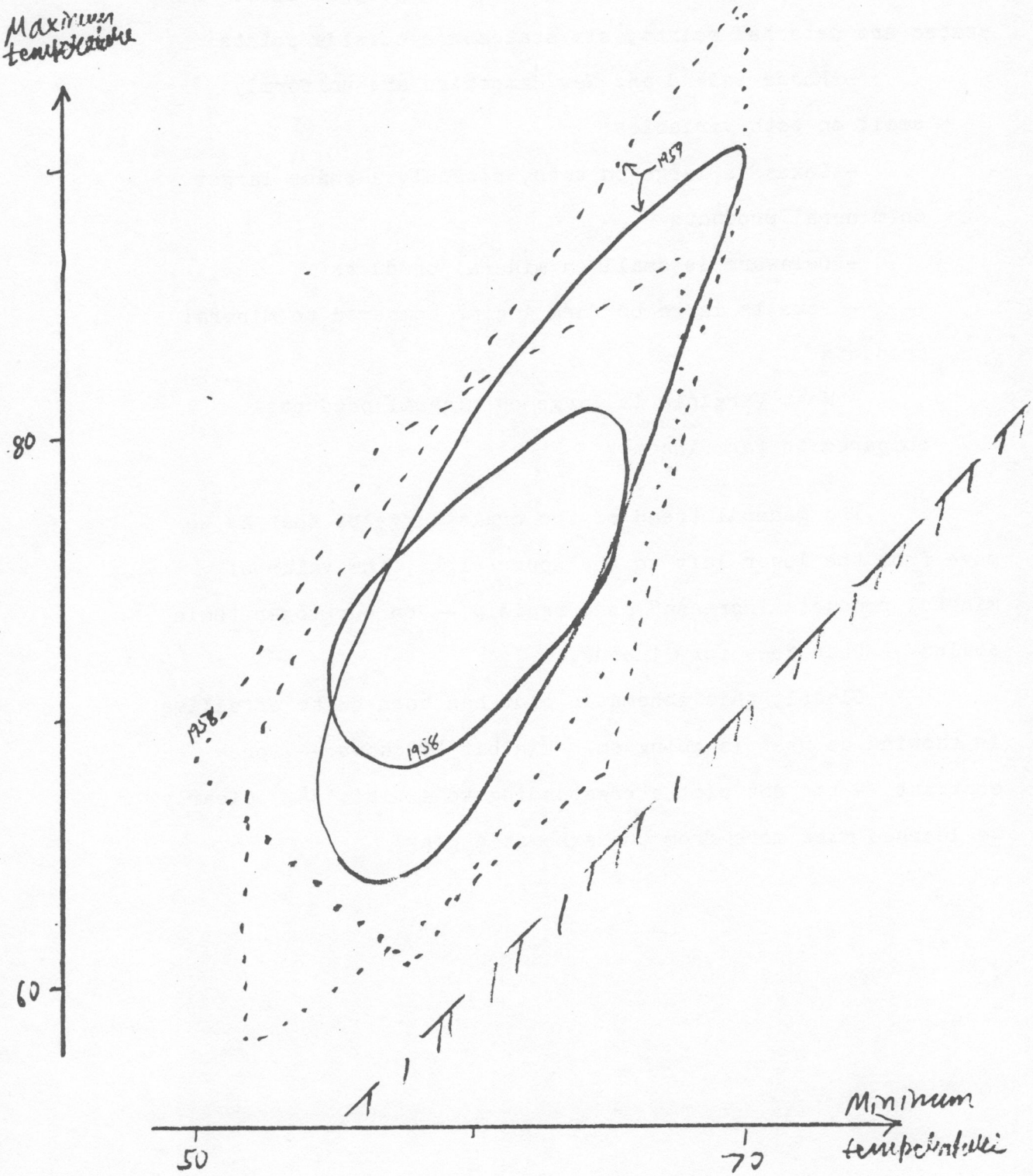


exhibit 24 of chapter 12

Two Junes in New York City



* an economic example *

Exhibit 25 gives farm income and value of mineral products for 48 states in 1958. Exhibit 26, using a logarithmic scale, gives the middle-oval-and-remaining-polygon plot. No states are detached points; six states are outside points:

- Rhode Island and New Hampshire are uniformly small on both variables
- Texas is large on both, possibly a shade larger on mineral products
- Delaware is small on mineral products
- Iowa is large on farm income compared to mineral products
- West Virginia is large on mineral products compared to farm income.

The general trend of the ovals suggests that as we move from the lower left to the upper right, the value of mineral products increases more rapidly -- on our logarithmic scales -- than does farm income.

Clearly this schematic plot has been quite effective in showing us what is going on. Exhibit 27 shows -- for contrast -- the dot plot corresponding to exhibit 34. Clearly we learned much more from the schematic plot.

exhibit 25 of chapter 12

Farm and Mineral Income Example
(further illustration of computation)

A) DATA, RAW AND LOGARITHMIC (millions of dollars)

	Farm income	Mineral production	x	y	x+y	x-y	2x-y
Ala	459.3	187.7	66	27	93	39	105
Ariz	394.5	314.5	60	50	110	10	70
Ark	599.9	131.6	78	12	90	66	144
Cal	2774.1	1502.7	144	118	262	26	170
Colo	566.5	395.3	76	60	136	16	92
Conn	165.9	13.1	22	-88	-66	110	132
Del	113.4	1.14	5	**-194**	**-189**	**199**	204
Fla	710.4	142.1	85	15	100	70	155
Ga	696.4	75.1	84	-12	72	96	180
Idaho	391.3	64.4	59	-19	40	78	137
Ill	2098.0	582.4	132	77	209	55	187
Ind	1167.9	197.6	107	30	137	77	184
Iowa	2563.8	85.4	141	-7	134	148	**289**
Kans	1114.8	498.5	105	70	175	35	140
Ky	563.9	402.1	75	60	135	15	90
La	359.6	1517.4	56	118	174	-52	4
Maine	203.9	12.6	31	-90	-59	121	152
Md	255.1	44.7	41	-35	6	76	117
Mass	168.7	23.9	23	-62	-39	85	108
Mich	722.9	343.5	86	54	142	32	118
Minn	1360.8	395.8	113	60	173	53	166
Miss	548.9	148.7	74	17	91	57	131
Mo	1081.5	144.0	103	16	119	87	190
Mont	472.8	177.2	67	25	92	42	109
Neb	1155.8	90.0	106	-5	101	111	217
Nev	47.9	68.3	-32	-17	-49	-15	-47
NH	62.9	3.88	-20	-141	**-161**	129	101
NJ	322.0	50.4	51	-30	21	81	132
NM	220.0	558.9	34	75	109	-41	-7
NY	857.3	204.9	93	31	124	62	155
NC	997.0	39.9	99	60	159	39	138
ND	673.8	59.1	83	-23	60	106	189
Ohio	1039.4	34.9	102	54	156	48	150
Okla	627.7	767.9	80	89	169	-9	71
Ore	391.4	45.0	59	-35	24	94	163
Penn	788.6	881.2	90	95	185	-5	85
RI	20.5	2.25	**-69**	**-165**	**-234**	96	27
SC	332.1	22.4	52	-65	-13	117	169
SD	648.5	41.5	81	-38	43	119	200
Tenn	486.0	124.9	69	10	59	59	128
Tex	2294.1	4038.7	132	161	**293**	-29	103
Utah	162.3	366.0	21	56	77	-35	-14
Vt	115.7	21.4	6	-67	-61	73	79
Wisc	457.2	203.2	66	31	97	85	101
Wash	556.2	60.9	75	-22	53	97	172
WVa	116.0	749.8	6	87	93	-81	**-75**
Wisc	1109.6	71.3	105	-15	90	120	225
Wyo	170.0	370.0	23	57	80	-34	-11

x = -200 + 100 log(farm income in millions of dollars)
y = -200 + 100 log(value of mineral production in millions of dollars)

(Cont'd)

exhibit 25 (Cont'd)

B) STEM-AND-LEAF DISPLAYS

for x

-6	9
-5	
-4	
-3	2
-2	0
-1	
-0	
0	566
1	
2	2313
3	14
4	1
5	96192
6	60796
7	86545
8	546301
9	390
10	753625
11	3
12	
13	22
14	41
15	

for y

-1	94,65,41
-9	0
-8	8
-7	
-6	257
-5	
-4	
-3	5058
-2	32
-1	2975
-0	75
0	
1	25760
2	75
3	011
4	
5	04467
6	0000
7	705
8	97
9	5
1	18,18,61,

for x+y

-2	34,
-1	89,61,
-0	
-6	61
-5	9
-4	9
-3	9
-2	
-1	3
-0	
0	6
1	
2	14
3	
4	03
5	93
6	0
7	27
8	0
9	3012730
10	019
11	09
12	4
13	6745
14	2
15	96
1	75,74,73,69,85
2	62,09,93

for x-y

-0	52,41,35,81,34
-2	9
-1	5
-0	95
0	
1	065
2	6
3	95295
4	28
5	5379
6	62
7	08763
8	571
9	6467
10	6
11	0179
12	190
1	99,48,

	T -94(-)	T -214(-)	T -282(-)	T -162(-)
	S -52(1)	S -132(2)	S -158(3)	S -97(-)
6h	E 6	E -66	E -54	E -22
48 12h	H 38	H -26	H 32	H 21
12h	H 96	H 60	H 136	H 96
6h	E 106	E 82	E 174	E 118
	S 164(-)	S 148(-)	S 278(1)	S 193(1)
	T 206(-)	T 230(-)	T 402(-)	T 258(-)
	dh 58	dh 66	dh 104	dh 75
	de 100	de 148	de 228	de 140

exhibit 25 (Cont'd)

for 2x-y

-0	47,07,14,75,11
+0	04,27,
5	
6	
7	019
8	5
9	20
10	589131
11	78
12	8
13	27128
14	40
15	5250
16	639
17	02
18	0749
19	0
2	04,89,17,00,25

T -158(-)
S -63(1)
6hE 16
48 12hH 91
12hH 170
6hE 190
S 267(1)
T 364(-)

dh 79
de 174

C) STATES OUTSIDE OR DETACHED

State	x	y	x+y	x-y	2x-y	*
Del	5	-194^{s}	-189^{s}	199	204	s
Iowa	141	-7	134	148	289^{s}	s
NH	-20	-141	-161^{s}	129	101	s
RI	-69^{s}	-165^{s}	-234^{s}	96	27	s
Tex	132	161	293^{s}	-29	103	s
W Va	6	87	93	-81	-75^{s}	s

D) SOURCES

World Almanac and Book of Facts for 1960, pages 660 and 700.
Their sources: Agricultural Marking Service and Bureau of Mines.

exhibit 26 of chapter 12

Values of Farm and Mineral Products
(logarithmic, by state, 1958)

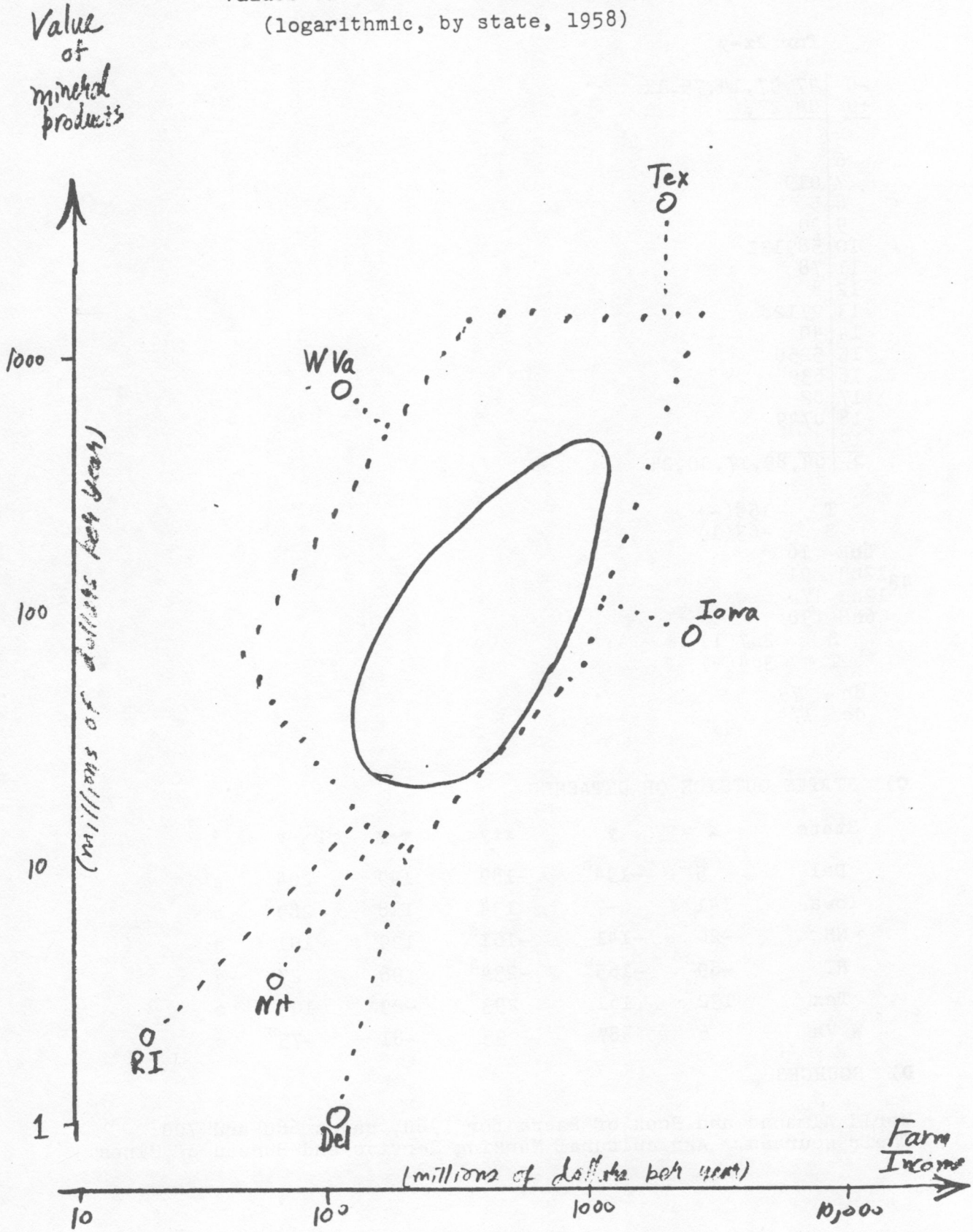
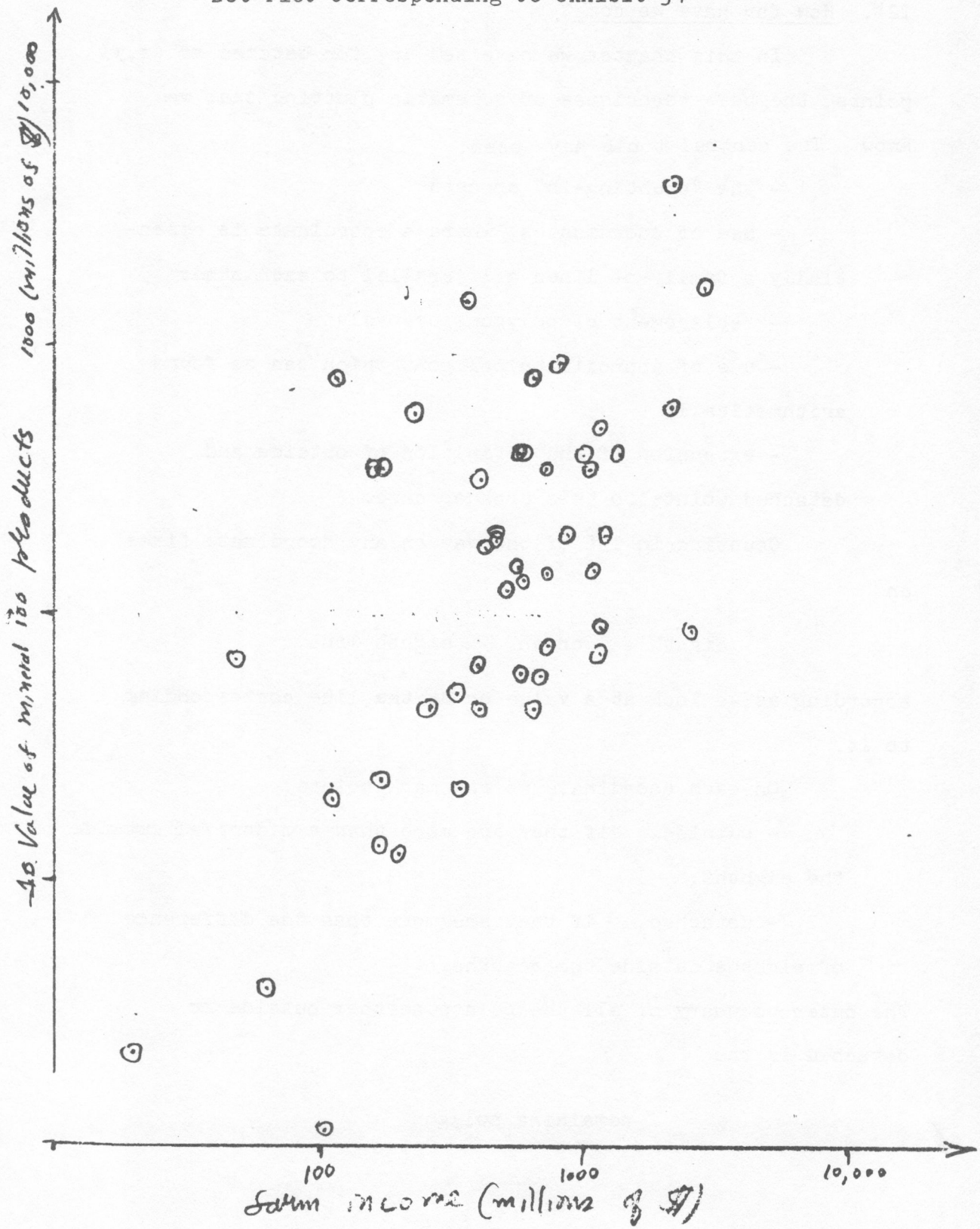


Exhibit 27 of chapter 12

Dot Plot Corresponding to exhibit 34



*****

12H. <u>How far have we come</u>?

In this chapter we have set up, for batches of (x,y) points, the best techniques of schematic plotting that we know. The central tools have been:

- the "counting-in" process
- use of coordinates, where a coordinate is essentially a family of lines all parallel to each other
- replacement of polygons by ovals
- use of approximate polygons which can be found arithmetically
- extension of the definition of outside and detached points to this broader case.

Counting in 1/8 of the way on any coordinate fixes an

eighth or an eighth line

according as we look at a value or at the line corresponding to it.

On each coordinate we can mark points:

- outside, if they are more than a midspread outside the eighths.
- detached, if they are more than the difference of eighths outside the eighths.

The outer boundary of all the points neither outside or detached is the

remaining polygon

To obtain a middle figure for a batch of points spread out on the plane, in principle, we count in 1/8 of the way according to every coordinate. In practice, we count in 1/8 of the way -- from both sides -- according to a few selected coordinates, thus fixing selected eighths and selected eighth lines. (We expect to include among these selected coordinates x, y, and some approximation to the popular diagonal of the rectangle formed by the eighth lines for x and y.) The polygon inside all the -- or all the selected -- eighth lines is a

middle polygon

The oval drawn -- by eye and hand -- inside this middle polygon is a

middle oval

Fairly often -- especially when confusion is easy -- the middle oval itself meets our needs for a schematic plot.

For some purposes, a middle oval combined with plotting of all the data points works well. Usually this happens only when there is only one batch and not too many points.

The combination of middle and oval (firmly solid) and remaining polygon (lightly dotted) often works quite well, especially when we are making comparisons in only slightly confusing circumstances.

When the situation is even simpler, we can often usefully:

- plot the outside points as open circles connected to the remaining polygon by dotted lines.

In the simplest and most confusion-free circumstances, we can also:

- plot the detached points as isolated filled circles.

(All these analyses and plots can also be carried through on an s-skipping or t-skipping basis, as section 12N, below, illustrates.)

*****

12K. The third dimension (background)

If we could plot -- and look at what we plotted -- in three dimensions, we could make a further major step. A coordinate would then be a family of planes, each parallel to the others. The result of counting-in from both sides along a coordinate would be a slice bounded by parallel planes. The common part of several slices would be a polyhedron. If we could find a middle polyhedron, we could probably fit a smoothish surface into it, thus getting to the equivalent of an oval.

Only one point of detail seems worthy of note: To find a middle polyhedron, we should act as if we were counting-in from 8 -- 8, surprisingly, not 6 -- independent directions. Thus we would use

sixteenths

not twelfths, in the place of eighths.

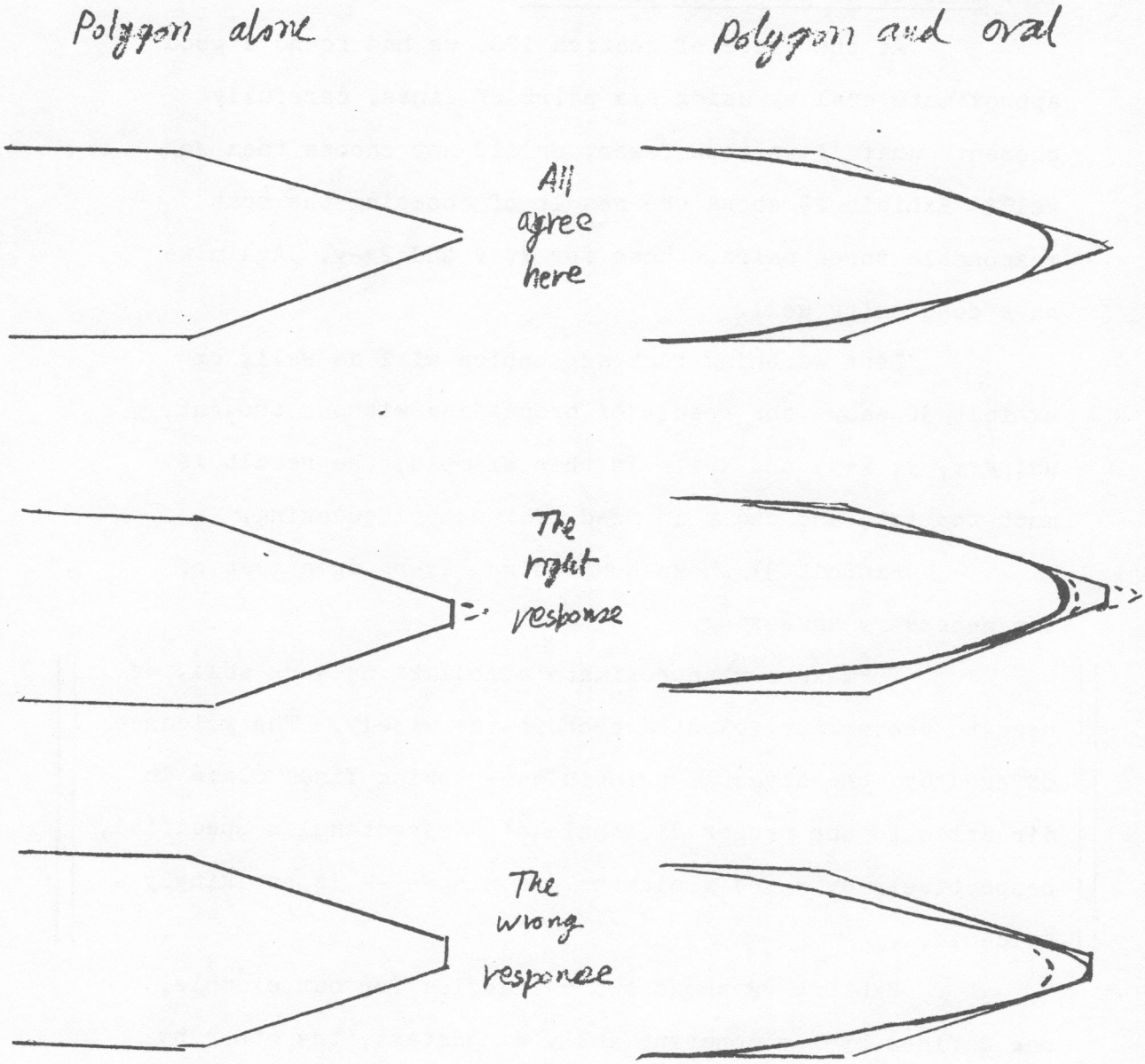
*****

12L. A word about drawing ovals (warning)

The simplest of the avoidable errors in darawing an oval is to be oversensitive to a little side at the end of a polygon. Exhibit 28 shows an illustration. In the top pair of sketches, we see a normally-drawn oval, one that penetrates reasonably far into the angle. If we cut off the top of the angle, as in the second and third lines, the right response is to shorten the oval a little, too -- as in the second line. The wrong response is to stretch the oval to touch the new side. As in so many other things, we need to practice moderation in drawing ovals.

exhibit 28 of chapter 12

Response to a Small Truncation



12M. <u>What if we pick less carefully</u>?

At the close of section 12C, we had found a good approximate oval by using six pairs of lines, carefully chosen. What if we took fewer, or did not choose them so well? Exhibit 29 shows the result of choosing the most reasonable three pairs, those for x, y and 2x-y. Again we have done quite well.

Lest we think that any choice will do well, ex exhibit 30 shows the result of proceeding without thought, using x, y, x+y, and x-y. In this example, the result is much too fat, and badly in need of further squeezing.

Exhibit 31 shows how adding 2x-y does most of the necessary squeezing.

To make our approximate calculations work well, we need to choose our selected coordinates wisely. The guidance offered by the diagonal principle -- taking lines close in direction to the proper diagonals of the rectangles specified, respectively by x and y eighths and hinges -- is strikingly valuable.

Exhibit 32 shows two rectangles for our example, one defined by x = constant and y = constant, the other by x+y = constant and x-y = constant. It is heartening to see how close they come to sharing a diagonal, since we plan to take eighth lines parallel to these diagonals next -- in either case.

exhibit 29 of chapter 12

June 1959 in New York City:
Comparison of Ovals

Solid oval: from x, y, 2x-y only

Dashed oval: from final polygon

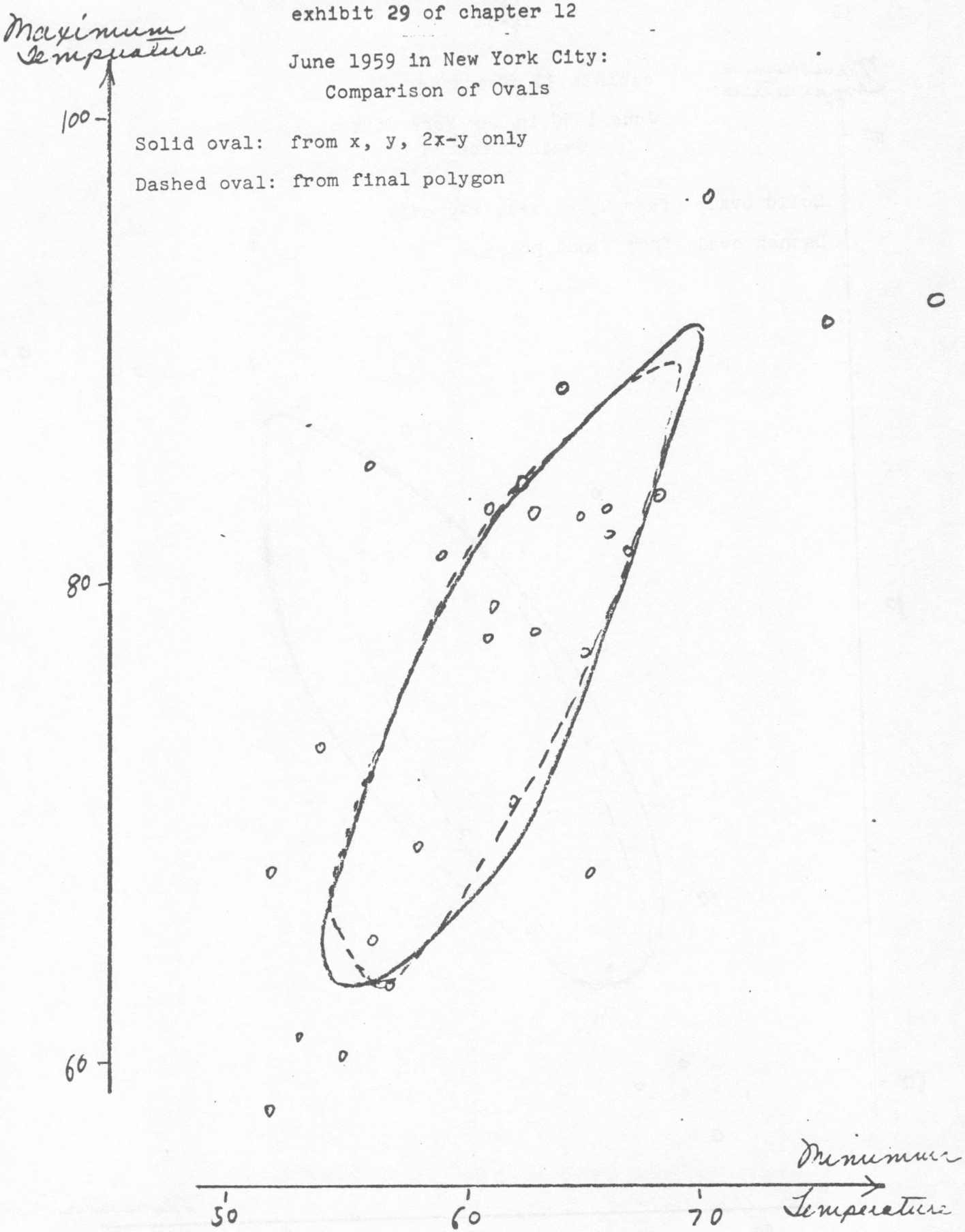


exhibit 30 of chapter 12

June 1959 in New York City:
Ovals Compared

Solid oval: from x, y, x+y, x-y only

Dashed oval: from final polygon

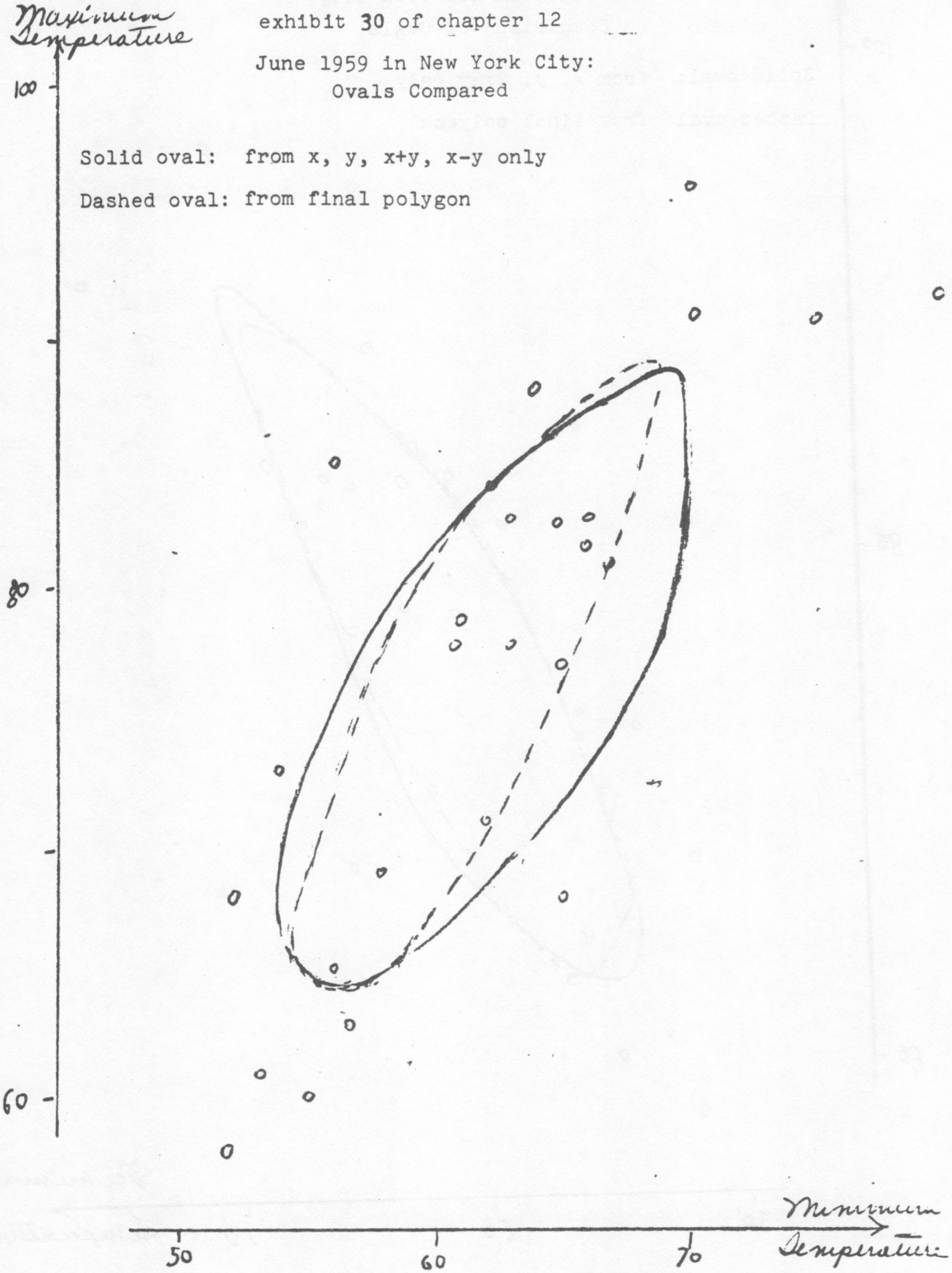


exhibit 31 of chapter 12

June 1959 in New York City:

Comparison of Ovals

Solid oval: from x,y,x+y,x-y,2x-y only

Dashed oval: final polygon

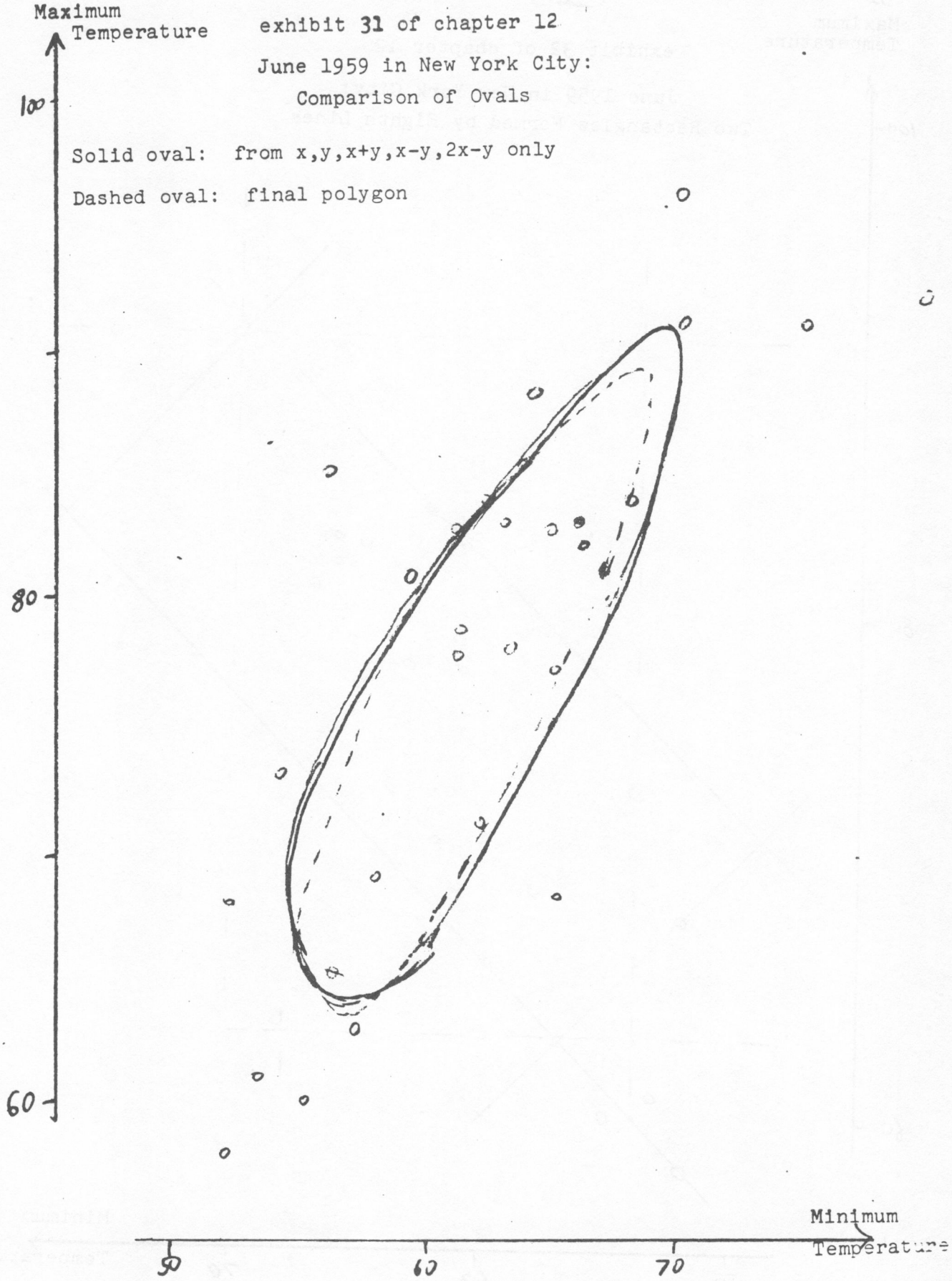
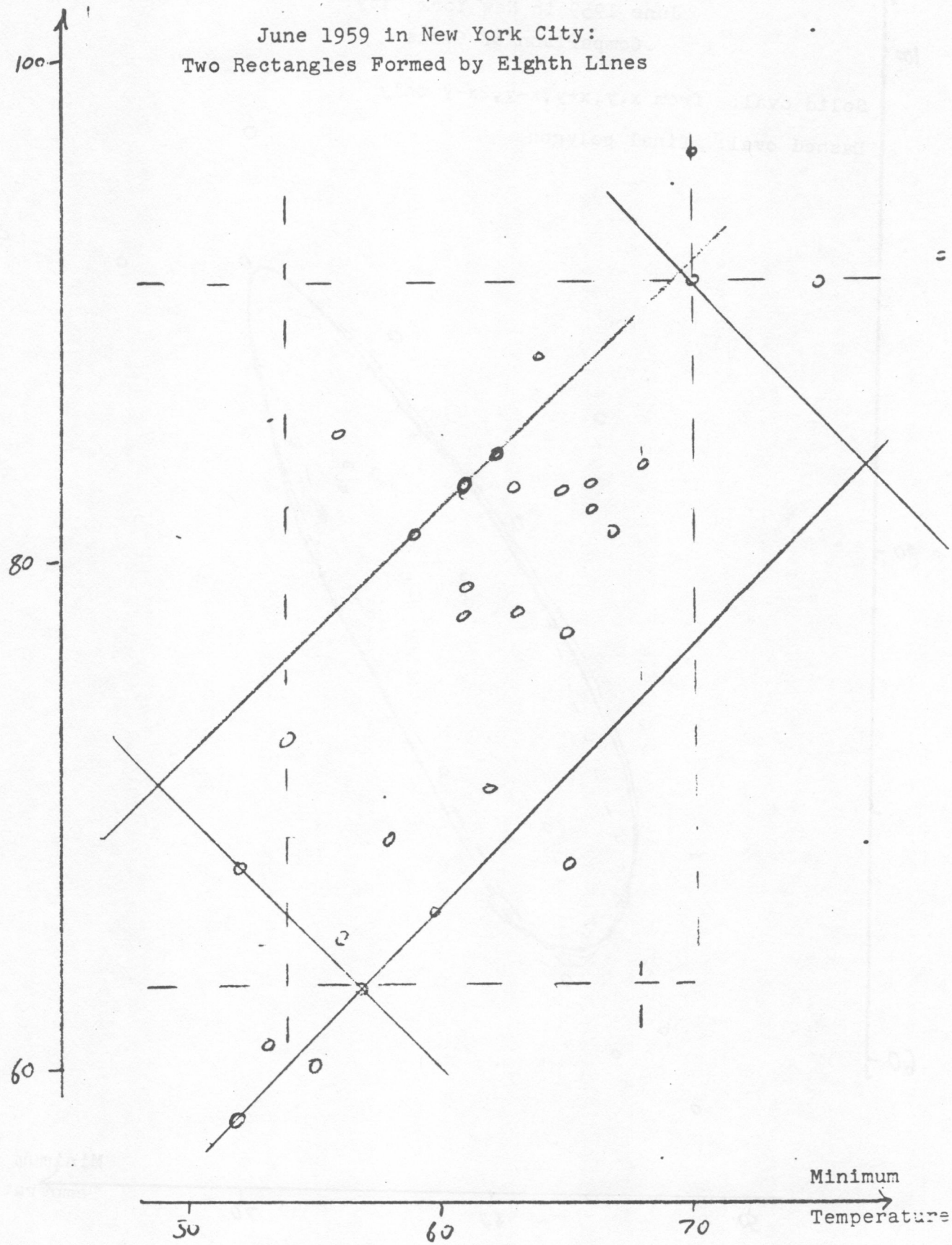


exhibit 32 of chapter 12

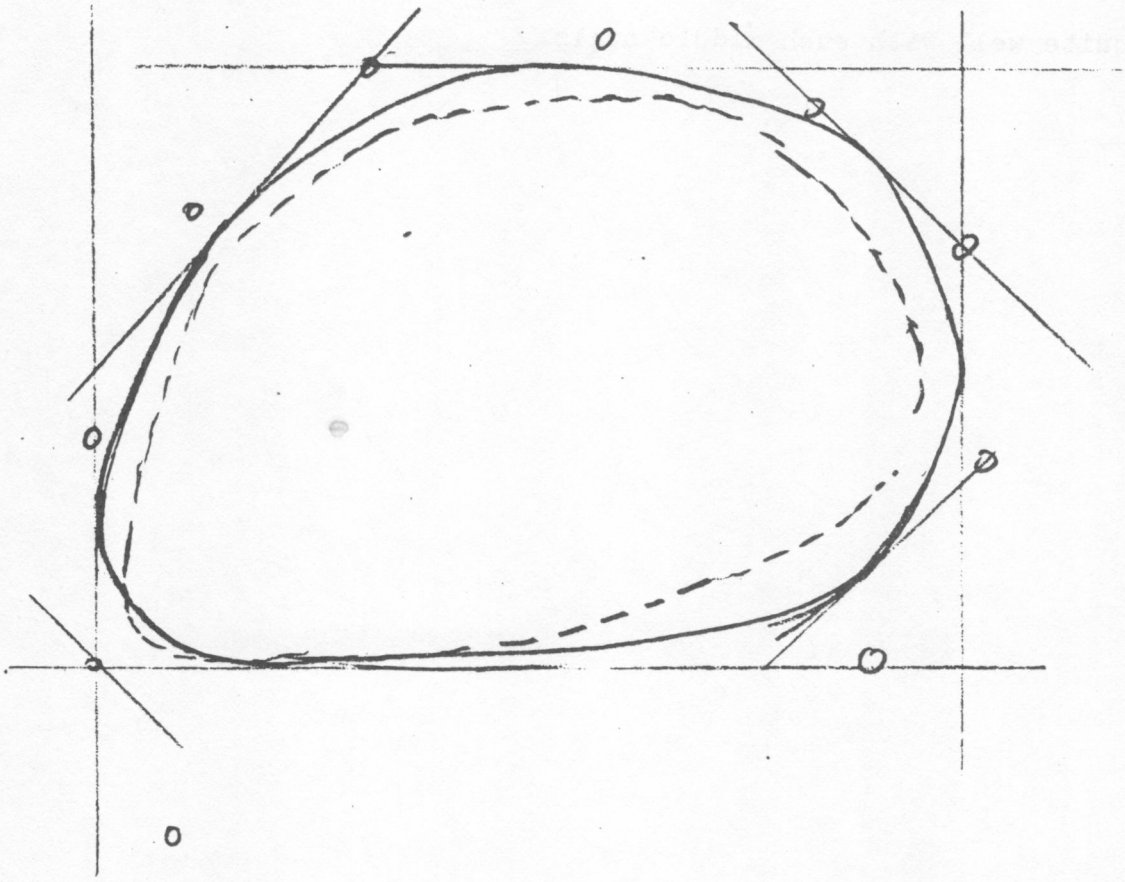
June 1959 in New York City:
Two Rectangles Formed by Eighth Lines



These approximation techniques are not perfect. We will occasionally do not quite so well. Exhibit 33 shows what x, y, x+y, and x-y will do for us in the somewhat pathological example of exhibits 7 and 8. The ovals do not match in detail, though we must admit that -- as quite schematic plots -- they give quite similar impressions. While we might have hoped for closer approximaion, we can get along quite well with such middle ovals.

exhibit 33 of chapter 12

Approximate Treatment of the Example of exhibit 7



Solid oval: From x, y, x+y, x-y, as shown

Dashed oval: From final polygon, see exhibit 8

12N. Skipping can be used (optional technique)

Once we have clear definitions of "outside" or "detached", we are ready to set aside points and make skipping analyses. Exhibit 34 continues our example; doing both S- and T-skipping. Only one other unusual point shows up, that for June 25th. (Why unusual days in June 1959 had round-number dates is a mystery -- and probably a happenchance.)

Clearly we can skip as part of our x-and-y together analysis whenever we wish.

In particular, we can use t- or s- middle ovals whenever we wish. Exhibit 35 compares u- and s- middle ovals for our example. (The remark of panel B of exhibit 34 has been used.) The difference between the two is appreciable rather than striking.

Exhibit 36 shows the schematic plot obtained by s-skipping. Comparing this with exhibit 25 shows us that the effect of s-skipping can be appreciable.

The farm and mineral income example of section 12F offers another opportunity for skipping. Exhibit 37 does the arithmetic; exhibit 38 shows the schematic plot. Of the six states outside in exhibit 26, three are detached in exhibit 38, as is Nevada. The other three originally outside in exhibit 26 are joined by Louisiana, New Mexico, Utah, and Wyoming.

exhibit 34 of chapter 12

Skipping analysis of June 1959 in New York City

(see exhibits 13 and 21)

A) S1 ANALYSIS -- eighths at 4 and hinges at 7 from 27 -- three points already set aside and omitted

for x

```
5|2234
5|56789
6|011122334
6|5556678
7|00
```

```
   T  41(-)
   S  46(-)
  4E 54
  7H 57
27
  7H 65
  4E 67
   S  75(-)
   T  80(-)
```

for y

```
5|8
6|013
6|56889
7|13
7|7889
8|112333344
8|8
9|1
9|6
```

```
   T   42(-)
   S   48(-)
  E 63
  H 68
  H 83
  E 84
   S   99(-)
   T  105(-)
```

for x+y

```
11|04
11|5
12|001
12|677
13|33
13|9
14|00124
14|668889
15|22
15|
16|1
16|6
```

```
   T    88(-)
   S    98(-)
  E 120
  H 126
  H 148
  E 152
   S   174(-)
   T   184(-)
```

for x-y

```
-2|6
-2|422210
-1|988776665
-1|421
-0|9986665
-0|3
```

```
   T   -38·
   S   -33·
  4E -22
  7H -20
27
  7H  -9
  4E  -6
   S     5·
   T    10·
```

for 2x-y

```
3|5679
4|003344
4|56777899
5|00123334
5|
6|2
```

```
   T  25 -
   S  31 -
  E 39
  H 43
  H 51
  E 53
   S  61 1
   T  67 -
```

for 3x-y

```
8|89
9|
9|688
10|02344
10|55568
11|1244
11|55689
12|00
12|7
```

```
   T    77 -
   S    85 -
  E  98
  H 102
  H 115
  E 119
   S   132 -
   T   140
```

exhibit 34 (cont'd.)

Notes: The stem-and-leaf displays are as in panel B of exhibit 13, less the entries covered by panel B of exhibit 21.

Since we have not included the three points set aside in the stem-and-leaf displays, we use depths in the 27 rather than in all 30.

B) OUTSIDE or DETACHED in A

Date	**x**	**y**	**x+y**	**x-y**	**2x-y**	**3x-y**	*****
25	65	68	133	-3	**62**^{s}	127	s

Remark: Setting aside this point (in addition to the three already set aside) would change a hinge for **2x-y** from 51 to 50 and an eighth for **3x-y** from 119 to 118. No further skipping would follow in an S1 = S analysis.

C) OUTSIDE or DETACHED -- after s-skipping

Date	**x**	**y**	**x+y**	**x-y**	**2x-y**	**3x-y**	*****
10	75	91	166	-16	59	**134**^{s}	s
20	56	85	141	-29	**27**^{s}	**83**^{s}	s
25	65	68	133	-3	**62**^{s}	127	s
30	**80**^{s}	92	172	-12	**68**^{t}	**148**^{t}	t

D) t-SKIPPING ANALYSIS

Problem: Carry out the t-skipping version of the same analysis.

exhibit 35 of chapter 12

u- and s- Middle Ovals Compared
June 1959 in New York City

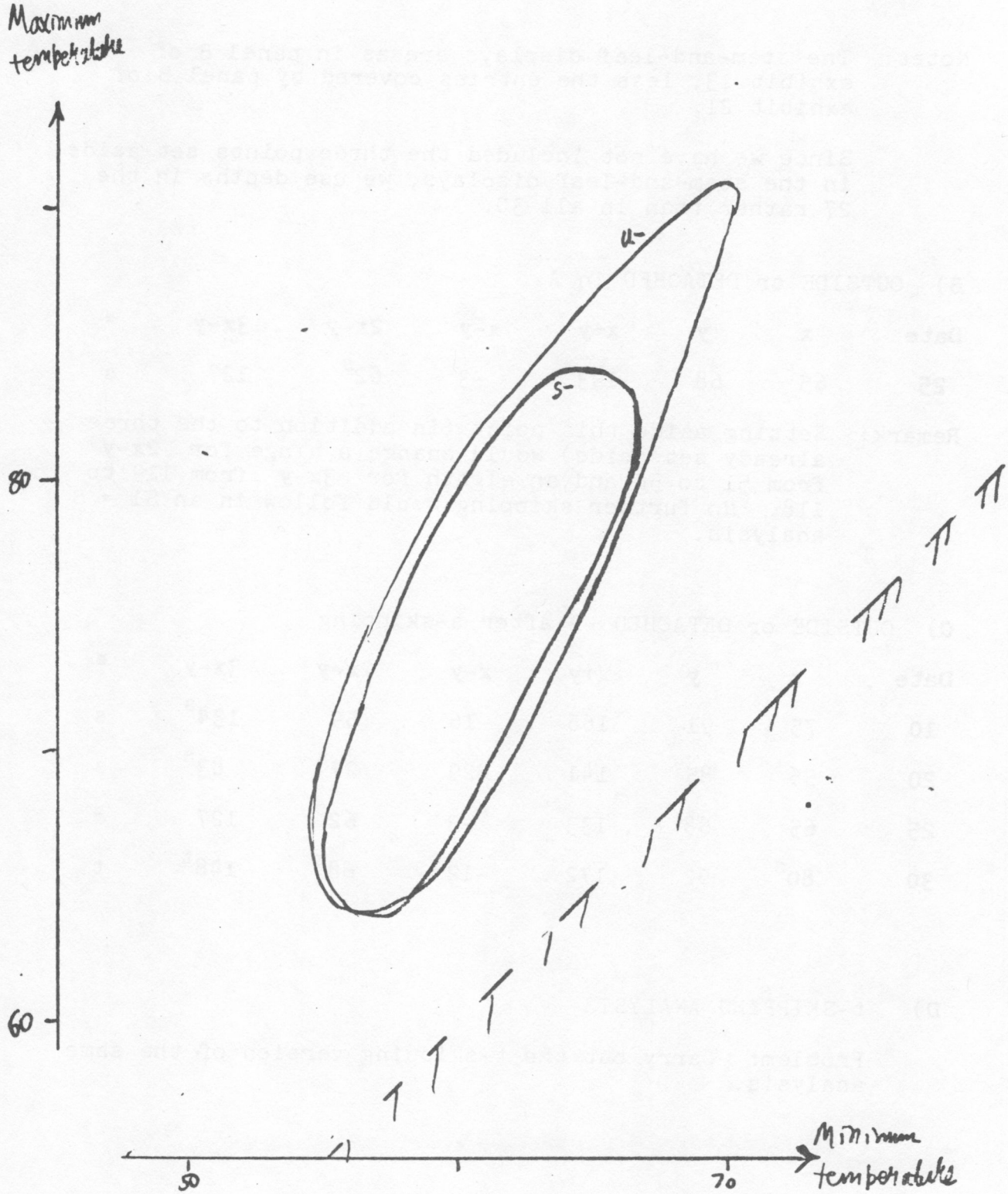


exhibit 36 of chapter 12

s-schematic plot for June 1959 in New York City

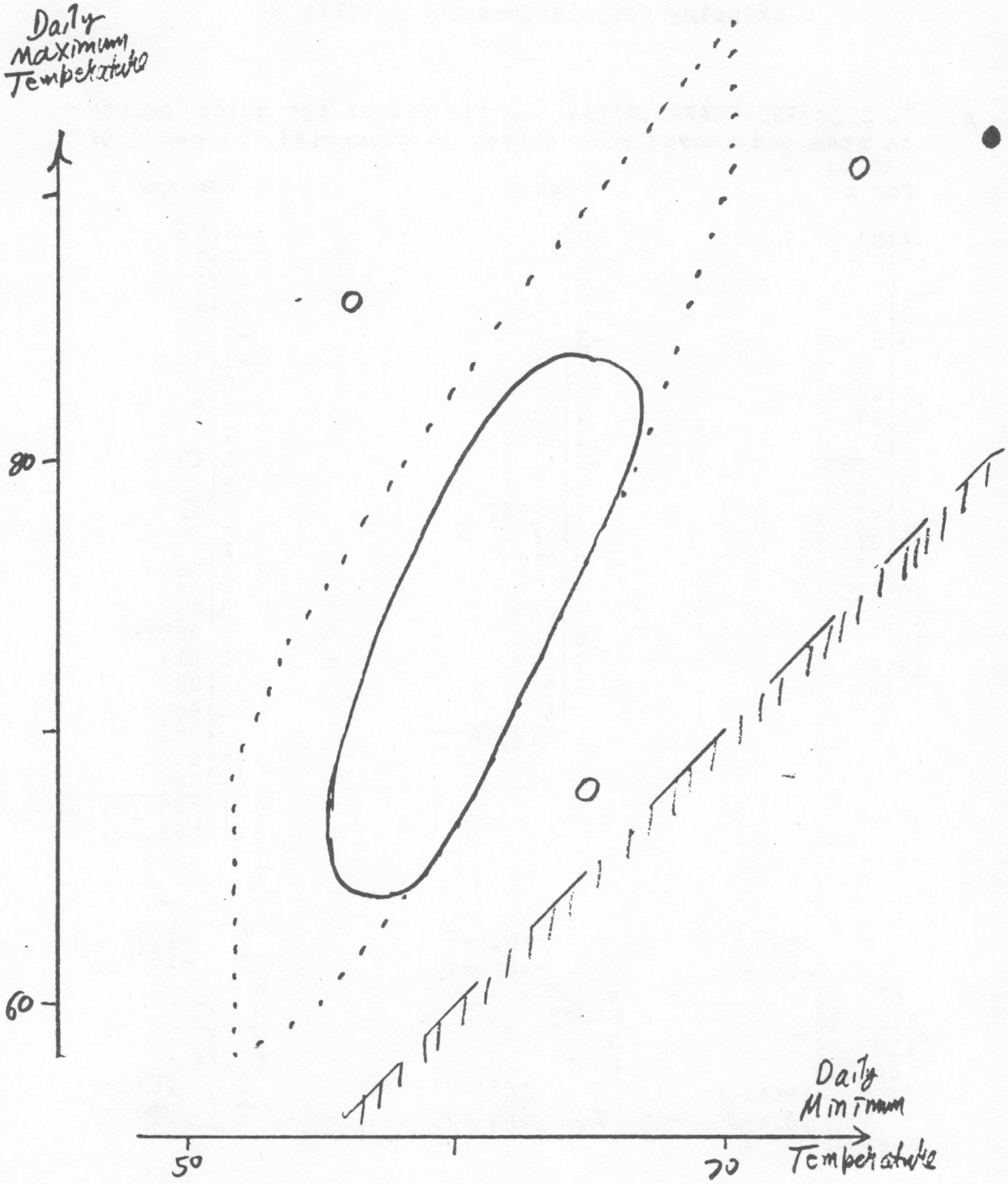


exhibit 37 of chapter 12

s-skipping Calculations for exhibit 33

A) S1 STEM-AND-LEAF DISPLAYS -- six states set aside, omitted in stem-and-leaves but counted in summaries, as per C of Exhibit 25.

for x

-3	2
-2	
-1	
-0	
0	6
1	
2	1233
3	14
4	1
5	12699
6	0679
7	45568
8	013456
9	039
10	235567
11	3
12	
13	2
14	4

for y

-9	0
-8	8
-7	
-6	752
-5	
-4	
-3	8550
-2	33
-1	9752
-0	5
0	
1	02567
2	57
3	011
4	
5	04467
6	0000
7	057
8	9
9	5
1	18,18

for x+y

-6	61
-5	9
-4	9
-3	9
-2	
-1	3
-0	
0	6
1	
2	14
3	
4	03
5	39
6	0
7	27
8	0
9	001237
10	019
11	09
12	4
13	567
14	0
15	69
1	75,74,73,69,85
2	62,09,

T	-59(-)	
S	-18(-)	
6E	23	
11H	52	
42 11H	93	
6E	105	
S	146(-)	
T	187(-)	

T	-151(-)	
S	-121(-)	
E	38	
H	-23	
H	60	
E	75	
S	158(-)	
T	188(-)	

T	-199(-)	
S	-107(-)	
E	-13	
H	43	
H	137	
E	173	
S	267(-)	
T	359(-)	

(Cont'd)

exhibit 37 (Cont'd)

for x-y

-0	52,41,35,34
-1	5
-0	95
0	
1	056
2	6
3	25599
4	28
5	3579
6	26
7	03678
8	157
9	467
10	6
11	0179
12	01

T -78(1)
S -32(4)
6E 16
42 11H 39
11H 87
6E 110
S 158(-)
T 204(-)

for 2x-y

-0	47,14,11,07
+0	04,
7	019
8	5
9	02
10	1589
11	78
12	8
13	12278
14	04
15	0255
16	369
17	02
18	0479
19	0
2	00,17,25

T -47(1)
S -7(3)
E 70
H 92
H 169
E 187
S 264(-)
T 304(-)

Note: " T " entries in parentheses are values outside in an earlier analysis -- here the u-analysis -- that are now detached.

B) ADDITIONAL STATES OUTSIDE IN **A**

State	x	y	x+y	x-y	2x-y
La	56	118	174	-52	4
NM	34	75	109	-41	-7
Nev	-32	-17	-49	-15	-47
Utah	21	56	77	-35	-14
Wyo	23	57	80	-34	-11

(Cont'd)

exhibit 37 (Cont'd)

C) S2 STEM-AND-LEAF DISPLAYS

for x

0	6
1	
2	23
3	1
4	1
5	1299
6	0679
7	45568
8	013456
9	039
10	235567
11	3
12	
13	2
14	4

for y

-9	0
-8	8
-7	
-6	752
-5	
-4	
-3	8550
-2	33
-1	952
-0	5
0	
1	02567
2	57
3	011
4	
5	044
6	0000
7	07
8	9
9	5
1	18,

for x+y

-6	61
-5	9
-4	
-3	9
-2	
-1	3
-0	
0	6
1	
2	14
3	
4	03
5	39
6	0
7	2
8	
9	001239
10	01
11	09
12	4
13	567
14	0
15	69
1	69,73,75,85,
2	09,62,

	T	-29(2)		T	-194(-)		T 199(-)
	S	2(-)		S	-139(-)		S 106(-)
5	E	41		E	-62		E -13
10	H	60		H	-23		H 43
37 10	H	99		H	54		H 136
5	E	106		E	70		E 173
	S	145(-)		S	147(-)		S 266(-)
	T	171(-)		T	202(-)		T 359(-)

(Cont'd)

exhibit 37 (Cont'd)

for x-y

-0	95
0	
1	056
2	6
3	25599
4	28
5	3579
6	26
7	03678
8	157
9	467
10	6
11	0179
12	01

for 2x-y

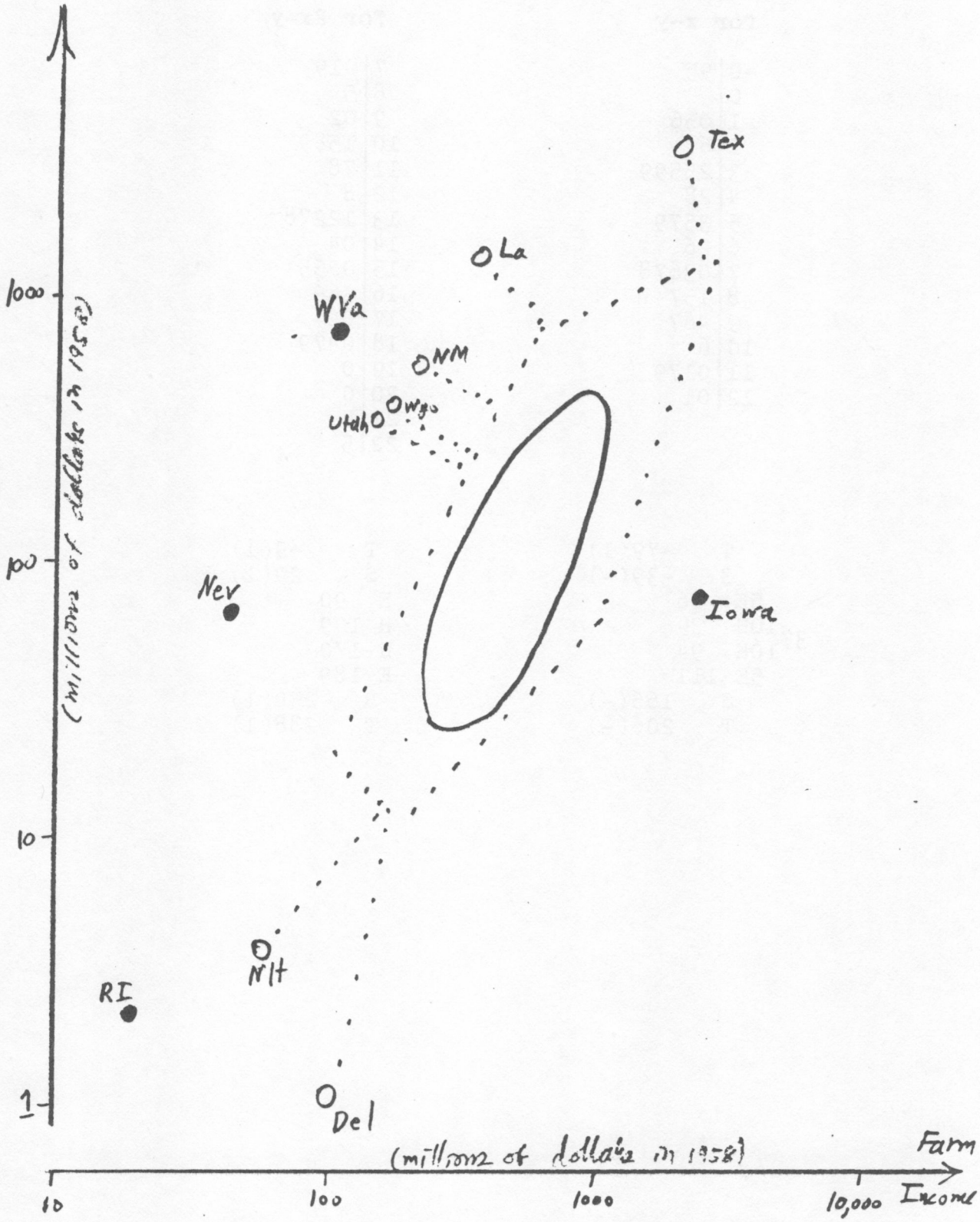
7	019
8	5
9	02
10	1589
11	78
12	8
13	12278
14	04
15	0255
16	369
17	02
18	0479
19	0
20	0
21	7
22	5

37

T -79(1)
S -39(-)
5E 16
10H 39
10H 94
5E 111
S 166(-)
T 206(-)

T -9(1)
S 29(1)
E 90
H 109
H 170
E 189
S 250(1)
T 288(1)

s-schematic Plot
Values of Farm and Mineral Products
(logarithmic both ways, by state)



All the newly distinctive states are high in mineral income as compared to farm income. (The unsymmetric behavior of the remaining polygon in exhibit 36 could have warned us to expect something of this sort.) All in all, the appearance of exhibit 38 seems more effective than that of exhibit 26.

* an intermediate form *

In this section we have played the full two-dimensional game, setting aside from each co-ordinate every point found to be outside on any coordinate. We can, if we like, use a weaker form of skipping, in which we set aside separately for each coordinate considered. This we call

coordinatewise skipping

Exhibit 39 shows the result for the 30 days of June, both in some detail and in a standard condensed form.

exhibit 39 of chapter 12

Coordinatewise skipping applied to the data of exhibits 13 and 21

A) ONE COORDINATE in DETAIL -- coordinatewise s-skipping

First analysis		Second analysis		
Summary	**Outside**	**Summary**	**Total outside**	
T 38	(-)	T 40	(-)	
S 45	(0)	S 46	(0)	
4E 54		4E 54		
30 8H 57		29 8H 57		
8H 66		9H 65		Stops
4E 70		5E 68		
S 79	(1)	S 76	(1)	
T 86	(-)	T 82	(-)	

B) MORE COORDINATES in STANDARD FORM -- coordinatewise s-skipping

	T	4 38(-)	4 60(-)	
	S	8 45(-)	8 46(-)	
	E	54	54	
	H	57	57	
for x	30 H	66	65	stops
	E	70	68	
	S	8 79(1)	9 76(1)	
	T	4 86(-)	5 82(-)	

	T	4 35(-)	
	S	8 48(-)	
	E	63	
	H	69	
for y	30 H	84	stops
	E	91	
	S	8 106(-)	
	T	4 109(-)	

(cont'd.)

exhibit 39 (cont'd.)

C) PROBLEMS

1) Describe the relation of the entries in panel B to those in panel A.

2) Complete panel B with standard forms for x+y, x-y, 2x-y, and 3x-y.

3) Compare the eighth polygons for this data: fully skipped, coordinatewise skipped, unskipped. (Three different polygons.)

*****

12P. Problems for chapter 12

1. Do the problem of exhibit 13.

2. Make a working plot based on panel B of exhibit 13 and problem 1, using x, y, x+y, 2x-y, and 4x-y. Draw the oval. Compare with exhibit 15.

3. Find maximum and minimum daily temperatures for another month (or the same month in another place). Carry through the arithmetic, draw the eighths polygon and the middle oval. [Members of a class can usefully make different choices.]

4. Complete problem 3 with a full schematic plot.

5. Find height and weight statistics for one or more of your favorite athletic teams. (Cover the whole squad if possible.) Carry through the arithmetic, draw the eighths polygon and the middle oval. [Members of a class can usefully choose different teams that play one another.]

6. Complete problem 5 with a full schematic plot.

7. Find the state-by-state Presidential vote for two successive elections. Using the two %'s Democratic, carry through the arithmetic, plot the eighths polygon, draw the middle oval. [Members of a class can usefully choose different pairs of adjacent elections.]

8. Complete problem 7 with a full schematic plot.

9. Find prices on two dates for 20 to 50 leading stocks. (You may want to select diversified stocks, or all from one industry, such as aircraft, oils, or banking.] Carry out the arithmetic, plot the eighths polygon, draw the middle oval. [Members of a class can usefully make different choices.]

10. Complete problem 9 with a full schematic plot.

11. Select a set of 20 to 50 pairs of numbers that interest you, carry out the arithmetic, plot the eighths polygon, draw the middle oval.

12. Complete problem 11 with a full schematic plot.

13. Do problem 1 of exhibit 39.

14. Do problem 2 of exhibit 39.

15. Do problem 3 of exhibit 39.

16. Relation of families in private residences to families with telephones about 1920. (22 Washington,669,430) translates as "In 1922 in Washington, D.C., 66.9% of families lived in private residences and 43.0% had telephones". Data-sets (61): (22 Washington,669,430), (22 Pittsburgh,611,374), 14 Baltimore, 686,164), (19 New Haven,246,245), (21 Portland,Me.,368,495), (15 Hartford,201,258), (16 Providence,266,265), (21,Springfield, Mass.,292,458), (20 Bridgeport,242,214), (17 Philadelphia,816,184), (22 Altoona,902,458), (20 Chicago,224,500), (21,Cleveland,453,324), (16 Evansville,890,346), (15 Grand Rapids,684,334), (21 Milwaukee, 400,396), (20 Indianapolis,844,538), (20 Akron,730,193),

(19 Detroit,499,306), (19 Youngstown,830,408), (20 Toledo, 766,420), (13 Montgomery,936,237), (13 Macon,946,218), (14 Charlotte,958,246), (16 Savannah,635,162), (20 Birmingham, 946,178), (15 Memphis,864,185), (20 Atlanta,833,286), (18 Mobile,924,197), (15 Chattanooga,892,237), (22 Richmond, 516,365), (19 Jacksonville,752,249), (20 Louisville,710,286), (16,New Orleans,853,116), (19 Tulsa,885,408), (21 Fort Worth, 920,373), (20 Little Rock,935,401), (21 Houston,878,443), (21 San Antonio,913,315), (21 Dallas,881,544), (16 Kansas City, 714,328), (17 St. Louis,353,250), (16 St. Joseph,900,410), (18 Oklahoma City,852,464), (21,Omaha,828,694), (18 Sioux City, 876,510), (16 Des Moines,858,488), (15 Duluth,620,520), (19 Lincoln,878,600), (21 Minneapolis,516,572), (14 Butte, 750,352), (16 Portland,808,490), (17 Denver,837,424), (18 Seattle,757,468), (18 San Diego,818,424), (17 Salt Lake City,860,512), (17 Los Angeles,784,473), (18 San Francisco, 531,474), (21 Spokane,860,570), (18 Sacramento,626,467), (21 Tacoma,878,463). Source: W. C. Helmle 1922. The relation between rents and incomes, and the distribution of rental values. Bell System Technical Journal 1 No. 2: 82-109. Table VIII on pages 103-104. Problem: Make a two-way schematic plot for all the cities given. Id tify outside and detached points. Comment on the results.

17. Using the same data, make 2-way plots on a single sheet for: (a) Cities for which data is for years 1913 to 1916; (b) Cities for which data is for years 1921 and 1922. Discuss the comparison. Make detailed 2-way plots on separate sheets. Do you see anything new?

18. Separate the same data geographically in a way making sense to you and make a comparative 2-way plot.

19. New method of determining the iodine numbers of oils and fats. (5 Royal,1708,1701) translates as "Iodine number of sample of 1945 crop, **Royal** variety by A.O.A.C. (official agricultural chemists) method was **170.8**, by new method was **170.1**". Data-sets for flaxseed oils (18): (5 Royal,1708,1701), (5 Redwing,1827,1834), (5 Dakota,1833,1831), (5 Bison,1685,1703), (5 Crystal,1871,1873), (5 Dominion, 1828,1823), (6 Royal,1830,1841), (6 Redwing 1912,1923), (6 Dakota,1909,1907), (6 Bison,1802,1816), (6 Crystal,1950,1955), (6 Dominion,1866,1869),(6 Gossamer,2003,2010), (6 Norfolk Queen,1930,1923), (6 Custer,1853,1868), (6 Victory,1904,1895), (6 Viking,1928,1928), (6 Novelty,1807,1813). Data-sets for soybean oils (16): (4 Early Blackeye,1269,1279), (4 Goldsoy, 1345,1349), (4 Manitoba Brown,1354,1367), (5 Pagoda,1330,1332), (5 Kabott,1317,1326), (5 Manitoba Brown,1354,1372), (5 Mandarin, 1307,1313), (5 La Pealon,1324,1309), (5 M-229,1308,1283), (6 Lincoln,1371,1375), (6 Mandarin,1283,1301), (6 Richland, 1261,1279), (6 Harman,1342,1343), (6 Le Platou,1381,1370),

(6 Kabott,1412,1398), (6 Pagoda,1412,1413). Data-sets for sunflower oils (7): (4 Sunrise,1277,1298), (4 3975,1345,1324), (4 3977,1251,1255), (4 3978,1303,1288), (5 Sunrise,1319,1325), (5 Mennonite,1360,1357), (6 Sunrise,1365,1382). Source: D. J. Hiscox 1948. Determination of iodine numbers. Analytical Chemistry 20: 679-680. Table 1 on page 679. Problem: Take as the two variables: "1/100 of the iodine number by the standard (A.O.A.C.) method" and "iodine number by new method minus iodine number by standard method". Make a 2-way schematic plot. What is the obvious way to change the plot to see better what is going on? Make that change. Summarize all you have been able to see. What do you think of the new method of measuring iodine numbers?

20. Class scores on different chapters of a statistics book. (1 computing instructions,-136,-179) translates as "On chapter **1, computing instructions**, the apparent difficulty was **-1.36** for 1952-3 and **-1.79** for 1954". (Negative scores mean low difficulty and thus high success by students.) Data-sets (18): (1 Computing instructions,-136,-179), (2 A taste experiment,-87,-99), (6 Normal distribution,-08,-99), (3 Binomial distribution,-47,-83), (7 Calculation of normal samples,-46,-56), (9 Comparison of two groups,-75,12), (10 Comparison of several groups,-39,26), (4 χ^2 distribution, -08,04), (5 Contingency tables,32,-34), (11 Simple experimental

designs,-20,38), (8 Internal estimation,26,-09), (15 Associated measurements,52,36), (16 Distribution of counts,41,56), (14 Log-ratio bioassays,126,04), (17 Meeting the assumptions of Anova,34,98), (13 Factorial experiments,71,167), (12 Regression,129,119). (The order of data-sets is roughly that of increasing difficulty.) Source: C. I. Bliss 1955. An introductory course in biometry for graduate students in biology. Biometrics 11: 213-224. Table 3 on page 222. Problem: Make a 2-way schematic plot. What can you see?

21. For the same data, use the change in score (from 1952-3 to 1954) and a suitable fraction of the sum of the two scores as x and y and make another 2-way schematic plot. How much recomputation did you do? How much did you need to do? Did you learn more from either plot than from the others?

22. New England haddock, changes in numbers in population with number caught. (32,462,9398) translates as "During 1932, the stock fell by 462 fish/day while 9,398 thousand fish were caught during the year." Data-sets (15): (32,462,9398), (33,2035,13829), (34,808,7044), (35,1014,8522), (36,1450,9041), (37,1311,11491), (38,855,7290), (39,905,8161), (40,910,11697), (41,1495,15817), (42,1063,13026), (43,1495,13708), (44,2404,20923), (45,1800,13246), (46,1618,12527). Source: H. A. Schuck 1949. Relationship of catch to changes in population size of New England haddock. Biometrics 5: 213-231. Table 8 on page 229.

Problem: Make an appropriate schematic plot. Tell what you see.

23. Quality of analysis of uranium tetrafluoride for uranium and fluorine. (2389,7602) translates as "Fluorine found 23.89%, uranium found 76.02%, hence +0.09% was unaccounted for". Data-sets (26): (2389,7602), (2400,7585), (1789,7961), (2209,7689), (2329,7623), (2262,7661), (2214,7675), (2223,7572), (2226,7578), (2364,7608), (2427,7570), (25,76), (20,75), (2279,7663), (81,60), (2408,7595), (2365,7610), (2441,7569), (24,72), (2282,7646), 2385,7577), (2411,7577), (2416,7677), (2416,7583), (2432,7575), (2428,7578). Source: J. O. Hibbits 1957. Precision of the pyrohydrolytic determination of fluoride and uranium in uranyl fluoride and uranium tetrafluoride. Analytical Chemistry 29: 1760-1762. Table III on page 1761. Problem: Make a 2-way schematic plot for % fluorine and % uranium. Comment on what you can see. Calculate the % unaccounted for for each data set. Make a 1-way schematic plot. Comment on what you can see. What 2-way schematic plot would you now recommend?

24. Yields of 25 varieties of alfalfa (called lucerne in some countries) in two succeeding years. (323,447) translates as "yield in 1949 was 3.23 tons per acre, yield in 1950 was 4.47 tons per acre. Data-sets (25): (323,447), (292,425), (358,415), (340,452), (354,466), (274,380), (278,410),

(314,376), (353,455), (351,458), (344,402), (368,486), (318,426), (362,406), (328,407), (368,371), (354,414), (346,426), (328,384), (337,391), (294,404), (316,387), (358,452), (340,366), (344,383). Source: R. G..D. Steel 1955. An analysis of perennial crop data. Biometrics 11: 261-212. Table 1 on page 203. Problem: Make 2-way schematic plot for the two year's yields. What can you see?

25. Friction between single rayon fibers after various finishing treatments. (None;327,189) translates as "After no finishing treatments, the coefficients of friction (averaged for 27 and 48 milligrams normal pressure) were $\mu_s = 0.327$ and $\mu_k = 0.189$". Data-sets (14): (None;327,189), (Vapoline oil,0.5%;373,300), (Vapoline 300%,1%;165,113), Vapoline 300%,0.5%,181,117), (Perma Par R,1%;156,093), (Velvamine K 90%,1%;179,108), (Veicol M 50,1%;220,130), (Veicol S,0.25%;355,165), (Spinol+triethanolamine,0.5%; 316,158), Varkaus,1%;355,230), (Sodium soap,0.5%;365,212), (Kerosene, direct application;334,167), (Pendulum oil, direct application;185,165), (Woolrex, direct application;298,232). Source: B. Olofsson and N. Graden 1950. Measurement of friction between single fibers. v. Frictional properties of viscose rayon staple fibers. Textile Research Journal 20: 467-476. Table 1B on page 468. Problem: Make a 2-way schematic plot. What can you see?

26. Cepheid variables; possible relation between their periods and the observed polarization of their light. (Polarization is acquired on the way here by the action of fine dust.) (SU Cas,290,405) translates as: "For the star **SU Cas** (Cas for Cassiopeia) the log of the period (in days?) is **0.290** and the amount of polarization is **0.0405** magnitudes". Data-sets (28): (SU Cas,290,405), (TU Cas,330,137), (DT Cyg, 398,60), (SZ Tau,498,215), (RT Aug,572,107), (SU Cyg,585,054), (α UMi,599,020), (T Vul,647,046), (FF Agl,650,125), (σ Cep, 730,100), (Y Sgr,761,059), (AW Per,810,314), (BB Sgr,829,413), (U Agl,846,579), (X Sgr,846,340), (η Agl,856,361), (RS Ori, 879,531), (W Sgr,880,184), (RX Cam,898,813), (W Gem,898,407), (S Sge,920,157), (ζ Gem,1006,023), (RX Aur,1065,374), (TT Agl,1138,149), (X Cyg,1215,106), (RU Cam,1345,050), (T Mon,1431,216), (SV Vul,1654,493). Source: T. Schmidt 1958. Interstellare Polarisation und Eigenfarben galaktischer Delta-Cephei-Sterne. Zeitschrift für Astrophysik 45: 214-238. Table 1 on page 220. Problem: Make a 2-way schematic plot. What can you see?

27. Cracks in aluminum alloy sheets, energy requirements to start a crack and to complete it (tearing the test strip apart). (2014-T6;10.0,13.3;7.6,10.6) translates as: "For a sheet 0.063 inch thick of alloy **2014** temper **6**, **10.0** inche-pounds were required to start a crack in a longitudinal direction and **13.3** inche-pounds in all were required to complete it.

In a transverse direction, the requirements were 7.6 inche-pounds to start and 10.6 inche-pounds to complete". Data-sets (14): (2014-T6;10.0,13.3;7.6,10.6), (2020-T6;3.4,4.8;4.2,3.2) (Alclad 2020-T6;3.2,2.2;2.5,1.8), (2024-T3;23.9,45.6;21.9,42.4), (Alclad 2024-T3;19.7,37.8;17.6,33.6), (2024-TS6;5.8,6.8;5.1,6.8), (5154-H38;20.9,38.2;23.2,39.2), (5456-0;33.4,78.4;31.4,76.0), (5456-H24;13.3,26.4;12.6,28.7), (6061-T6;18.6,52.2;16.6,40.1), (6071-T6;8.5,17.0;9.4,13.0), (7075-T6;12.0,17.0;8.7,12.4), (7079-T6;22.7,34.4;15.1,28.6), (7178-T6;6.4,6.4;5.6,6.5).
Source: J. G. Kaufman and A. H. Knoll 1963. Kahn-type tear tests and crack toughness of alluminum alloy sheet. Materials Research and Standards 4:151-155. Table 1 on page 152.
Problem: Make a 2-way schematic plot of the starting and total energies in the longitudinal direction. What can you see?

28. For the same data-sets, make a similar plot in the transverse direction.

29. For the same data-sets, make a 2-way plot with differences between total and starting energies in the longitudinal and transverse directions as horizontal and vertical coordinates.

30. Find, from other sources, at least two batches of (x,y) data-sets that interest you and make a 2-way schematic plot for each.